

November 2005

Published_Applications Nucleic Acid and Published_Applications Amino Acid database searches now generate two sets of results each. The Published_Applications databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Newly published applications will appear in the Published_Applications_New databases; older published applications make up the Published_Applications_Main databases.

Searches run against Nucleic Acid Published_Applications produce two sets of results, with the extensions **.rnpbm** (Published_Applications_NA_Main) and **.rnpbn** (Published_Applications_NA_New). Searches run against Amino Acid Published_Applications produce two sets of results, with the extensions **.rapbm** (Published_Applications_AA_Main) and **.rapbn** (Published_Applications_AA_New).

THIS PAGE BLANK (USPTO)

Db 400 EPEAELEERAKSAKPP--QSSSTEPARKP-CQKEKRVPEEKQQA-KPVKVERT-RKRS 454
QY 272 EAAAAKEKNSKQSRVSTD 289
Db 455 EGFSDMRKVEKKEPSVE 472

RESULT 2

US-11-135-855-29
; Sequence 29, Application US/11135855
; Publication No. US20050255557A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; APPLICANT: SMITHKLINE BEECHAM P.L.C.
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50013
; CURRENT APPLICATION NUMBER: US/11/135,855
; CURRENT FILING DATE: 2005-05-24
; PRIOR APPLICATION NUMBER: US/10/203,708
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/04703
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/182,172
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 60/186,084
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 717
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-135-855-29

Query Match 6.3%; Score 143.5; DB 7; Length 717;
Best Local Similarity 24.8%; Pred. No. 0.003;
Matches 79; Conservative 56; Mismatches 138; Indels 45; Gaps 15;

QY 2 AFQSSPLVKFASPAESFLS-APGDNFTSLFADSTP-----STLNPRDMWTPD 49
Db 211 ARKASSLDQASVSSEENSESSSEKTSQDFTPEKKAAVRAPRRGPLGGRKKKKAP 270
QY 50 SVADIDSRL---SVIPESQDAEDDESHSTSATPSTSE---KKPVKKRKSQGVLPPEPKT 103
Db 271 SASDSKADSGAKPEPVAMARSASSSSSSSSSDSDSVVKKPPGRKPAEKPLPKPRG 330
QY 104 NLPPKR--AKTEDEKEQRRVRLNRRRAAQSRRERKLEVEALEKRNKELETLLINQV 161
Db 331 RKPKEPSPSSSSSDSDSDVDRISEWKR---RDEARRRELEARRRRQEBELRLRQ 386
QY 162 KTNLLILVEELNRRSSGVVTRSS--SPLDL--QDSITLSQQLFGSRDQ-----TWS 211
Db 387 EKE--EKERRERADRGEAERGSSGDELREDDEPVKKR---GRKGRGPPSSSDS 440
QY 212 NPEQSLMDQIMRSAANPTVNPASLSPSPPIISDKFQTKEEDEQADEEMEQTWHETK 271
Db 441 EPEAELEERAKSAKPP--QSSSTEPARKP-CQKEKRVPEEKQQA-KPVKVERT-RKRS 495
QY 272 EAAAAKEKNSKQSRVSTD 289
Db 496 EGFSDMRKVEKKEPSVE 513

RESULT 3

US-10-131-826A-160
; Sequence 160, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 160
; LENGTH: 605
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-131-826A-160

Query Match 6.0%; Score 136.5; DB 1; Length 605;
Best Local Similarity 20.2%; Pred. No. 0.0073;
Matches 104; Conservative 62; Mismatches 171; Indels 177; Gaps 22;
QY 7 SPLVKFEASPAESFLSA-----PGD---NFTSLFADSTPSTLNPRDMWTPDSVADI-- 54
Db 15 SPLLLLLVGG--FLGACVAGSDEFGPEGLTSTLLDLLPTGLEPLDSEEPSETWGLGA 72
QY 55 -----DSRLSVIP-----ESQDAEDDESHSTSATA-----PSTSEKKPVKKR 91
Db 73 GLGASGSGFSPSENEESRIQLPQQVFWEEELNDSLDLPTADYVFPDLTEKAGSIED 132
QY 92 KSWGQVLP-----EPTNLPKPKRAKTEDEKEQRRVRLNRRRAAQSRRERK 139
Db 133 TSQAQELPNLPSPLPKMNLVPEPPHMPPEEEEEEEEEER----- 173
QY 140 RLEVEALEKRNKELETLLINQVKTNLILVEELNRRSSGVVTRSSSPLDSLQDSITLSQ 199
Db 174 --EKEVEKQEEEEEEELLPVNGSQ-----EAKPQVRDFSLTSSSQTP----- 215
QY 200 QLFGSRDGTMSNPQSLMDQIMRSAANPTVNPASLSPSPPIISDKFQTKEEDEQADE 259
Db 216 -----GATKSRHEDSGQASSGVVESSMGPSLLPLPSVTPPT-----VTPGDQ 258
QY 260 DEEMEQTWHETKEAAAAAEEKNSKQSRVSTDSTORPAVISGGDAAVP-VFSDDAGANCLGL 318
Db 259 DSTSQE-----AEATVLPAAGLGVFEAFQEAASEATAGAAGL 296


```
QY 319 DPVHDDGPPSIGHSFGLSAAADRYLLESQLLASPNASTVDDVYLAGDSAACTNPL- 377
DB 297 SQHEEVPALP--SF-----POTAPSGAEHPDEDEPLGSR TSA--SSPLA 337
QY 378 PSYDVF-----DINFLTDANHAA-----VDIVAASNYAAAD-- 410
DB 338 PGDMELTPSSATILGOEDLNQQLLEGQAAEQRIPWDSTOVICKDSNLAGKRYIILNMT 397
QY 411 RELDLEIHDENQIPSRHSIQQ---PQSGASSHG 441
DB 398 ENIDCEVF--ROHRGPOLLALVEEVLPRHSGCHG 430

RESULT 4
US-10-485-517-129
; Sequence 129, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WQ
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 129
; LENGTH: 895
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-129

Query Match 5.4%; Score 124.5; DB 1; Length 895;
Best Local Similarity 22.2%; Pred. No. 0.076;
Matches 98; Conservative 63; Mismatches 175; Indels 105; Gaps 21;

QY 55 DSRLSVIPESQDAEDDESHTSATATSEKPKVKRKSQGWGQ-----VLPEPKT 103
DB 405 DKKLPELVSYDSKDYAYIRFPVPSNGTREVKTIVSSIE-YGENIHBDYDYLTVFAQPIT 463
QY 104 NLP-----PRKAKTEDEK--EQRVERVLRNRAAQSSRRERKLEVEA 145
DB 464 NNPDDVDEETYNLQKLLAPYHKAKTLERQVYELEKLEKLPKYKAEYKKKLDQTRVE- 522
QY 146 LEKRKKELETLLINVQKTNLILVEELNRRFRSSGVVTRSSPLDSLQDSITLSQQLFGSR 205
DB 523 LADQVKSATVEFENVPTN---DQLTDLQEAHFVVFESEENSESVMDGVFHPFFYATL 578
QY 206 DQG---TMENPEQSL-MQIMRSAANPTV--NPASLSPSL--PPISDK----- 245
DB 579 NGQYVVMKTKDSDYKWLIVGKRVTVTSKOPKNSRTLIFPIYIPDKAVYNAIVKVVVA 638
QY 246 -----EFQKKEDEEQADEEME-QTWHETKEAAAKKNSQKSRVSTD 289
DB 639 NIGYEQYHVRINQDINTKDDTSQNTSEPLNVGTQEGKVADTVVAENSTATNPKD 698
QY 290 STORPAVSIIGGDAAVPVFSD---DAGANCLGLDPVHVHDDGPPSIGHSFGLSAAALDADRYL 346
DB 699 ASDKADV-----IEPESDVVKDADNN---IDKDVQHD-----VDHLSDMSDNNHFDKYD 744
QY 347 LESQ--LLASPNASTVDDVYLAGDSAACTNPLPSYDFDIN-----DFLTDANH-- 395
DB 745 LKEMDTQIAKDTORNVDKO---ADNSVGMSSNVDTDKDSNKNKDKVQLNHIADKNHGTG 801
QY 396 --AAYDIVAASNYAAADRELD 414
```

```
DB 802 KKAALDVV-KQNYNNTDKVTD 821

RESULT 5
US-10-821-234-1120
; Sequence 1120, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1120
; LENGTH: 4384
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1120

Query Match 5.4%; Score 123.5; DB 1; Length 4384;
Best Local Similarity 22.2%; Pred. No. 0.61;
Matches 100; Conservative 58; Mismatches 202; Indels 91; Gaps 22;

QY 53 DIDSRLSVIPESQD-AEDDESHTSATATSEKPKVKRKSQGWQLPEPKTNLPPKRA 111
DB 2996 ELSQKLSSQSSMSKETVETQHFNSIEDKVTYSEISKVSHQSVGL-----CPPLEET 3048
QY 112 KTEDEKEQRRVERVLRNRAAQSSRRERKLEVEALEK-----RNKELETLLINVQKT 163
DB 3049 ETSPTKSPDSLE--FSPGKESPSDDVFDHSPIDGLEKLAPLAQTEGKEIKTLPVYVSFV 3106
QY 164 NL-----ILVEELNRRFRSSGV--VTRSS-----SPLDSLQDSITLSQQLF 202
DB 3107 QVGKQVEKEIQGGVKKIISQECKTVQETRGTPYTTTQOKQPPSPQSGPDD--TLEQVSF 3165
QY 203 GSRDQGTNSNPEQSLMDQIMRSAANPTVNPASLSPSLP--PISDKKEFQTKKEDEEQADE 260
DB 3166 LBSGKSLPTPTPSESEVSEYFTSKT--PDSLIAVIPKSPDPIPEVSESEEEQA-KS 3222
QY 261 EMEQOTWHETKEAAAKKNSQKSRVSTDSTORPAVSIIGGDAAVPVFSDDAGANCLGLDP 320
DB 3223 TSLKQT--TVEETAVERE---MPNDVSKDSNQRPK---NNRVAYIEFPFPPPP---LDADQ 3271
QY 321 VHODGPPPSIGHSGFGLSAAALDA-----DRYLESQLLASPNASTVDDVYLAGDSA- 370
DB 3272 IESDK-----KHHYLPKEKVDIMI EVNLQDEHDKYQLAEFPVIRVQPPSPVPPPGADVSDSD 3326
QY 371 -ACFTNPLP-SYDVEDINDFLTDD-----ANHAAYDIVAASNYAAADRELDLEIHD 419
DB 3327 DESIYQVPVKKYTKFLKE--VDDEQEKPKASAEKASNQKELESNGKDNFGLGLDS 3384
QY 420 PENQIPSRHSIQQPOSGAS-----SHGCD 443
DB 3385 PQNEIAQNGNDQSI TECSIAATTAEFSDHTD 3415

RESULT 6
US-10-485-517-212
; Sequence 212, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
```

[illegible]

```

: Sequence 252, Application US/10485517
: Publication No. US20050256299A1
: GENERAL INFORMATION:
: APPLICANT: University of Sheffield
: APPLICANT: Biosynexus Incorporated
: APPLICANT: Foster, Simon
: APPLICANT: Mond, James
: TITLE OF INVENTION: Antigenic Polypeptide
: FILE REFERENCE: F100629WO
: CURRENT APPLICATION NUMBER: US/10/485517
: CURRENT FILING DATE: 2004-02-02
: PRIOR APPLICATION NUMBER: GB 0118825
: PRIOR FILING DATE: 2001-08-02
: PRIOR APPLICATION NUMBER: GB 0200349
: PRIOR FILING DATE: 2002-01-09
: NUMBER OF SEQ ID NOS: 424
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 252
: LENGTH: 761
: TYPE: PRT
: ORGANISM: Staphylococcus aureus
: US-10-485-517-252

```

Query Match	5.2%	Score 118.5;	DB 1;	Length 761;
Best Local Similarity	18.4%	Pred. NO. 0.16;		
Matches 87;	Conservative	72;	Mismatches 204;	Indels 111; Gaps 13;

```

Qy      4 QSSPLVKPEASPAESFISAPGDNFTSLFADOSTSTLTNPRDM----- 46
    :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db     189 ESKSDTSMISMOSQTSGTSTSTSLSDSTSTLSLGSAMNQSVDSNSAQSN 248
    :   :   :   :   :   :   :   :   :   :   :   :   :
Qy     47 TPDVSADDSRLSLVIPESQDAEDDESHSTSATASTSEKKPVKRSW-CQVLPEPTNL 105
    :   :   :   :   :   :   :   :   :   :   :   :   :
Db    249 TSTSTSEDQSSTSYTQSTSQSGETSTST--SLSDTSISKSTSQSGVSTSA 304
    :   :   :   :   :   :   :   :   :   :   :   :   :
Qy    106 PPKRAKTEDEQRVRVILNRRAAQQRRKRLEVALEKENKELETLLINVOKTNL 165
    :   :   :   :   :   :   :   :   :   :   :   :   :
Db   305 -----GSESSEDSQSISTSAESTSBASSTSLSDSTSTSN 339
    :   :   :   :   :   :   :   :   :   :   :   :   :
Qy    166 ILVEELNRPRSSGVVTRSSPDLSDQSIYTLSQLFGSRDQGTMSNPEOGLMDQMRA 225
    :   :   :   :   :   :   :   :   :   :   :   :   :
Db   340 SGASTSTSLGNASAGASEDLSTSLSDTSASMOSSES-DSQSTS--ASLSDSLST 395
    :   :   :   :   :   :   :   :   :   :   :   :   :
Qy    226 ANPTVNPAISGPSLPPIIDKFQTKEEDEQADEMEQTHETKEAAAAAKENSKQRS 285
    :   :   :   :   :   :   :   :   :   :   :   :   :
Db   396 SNRMSTIASLSTSV-----STSESGSTSESTSES 426
    :   :   :   :   :   :   :   :   :   :   :   :   :
Qy    286 VGT-----DSTORPAVSTGGAAVPVFEDDAGANCLGLDPHQDDGFPSIGHSGFLGAAL 340
    :   :   :   :   :   :   :   :   :   :   :   :   :
Db   427 TSTSLSDSQSTR-STASGSGASTSTSTSDSRST-----SASTSTM 467
    :   :   :   :   :   :   :   :   :   :   :   :   :
Qy    341 DADRVLLESOLLASNASTVDVDDLAGDSACAFNTNPLPS--DYDFINDFLT-D--DANH 395
    :   :   :   :   :   :   :   :   :   :   :   :   :
Db   468 RPTSDDSQGMSLSTSTSTSMGSDSTSLSDSVSDTSDSTASTSGSMVSVLSLSDSTST 527
    :   :   :   :   :   :   :   :   :   :   :   :   :
Qy    396 AAYDIVAASNYAADRELDLFIHPDENQIPSRHSIQOPQSGASHGCDGGIAV 449
    :   :   :   :   :   :   :   :   :   :   :   :   :
Db   528 SASEVWSAS--ISDSQSMSESWNDSESVSENBSDESCKMSGSTSVSDSGSLSV 579
    :   :   :   :   :   :   :   :   :   :   :   :   :

```

RESULT 10
US-10-984-645-2
; Sequence 2, Application US/10984645
; Publication No. US20050244386A1
; GENERAL INFORMATION:
; APPLICANT: Habener, Joel
; APPLICANT: Zulewski, Hendrik
; APPLICANT: Abraham, Elizabeth
; APPLICANT: Vallejo, Mario
; TITLE OF INVENTION: METHOD OF TRANSPLANTING IN A MAMMAL AND TREATING DIABETES MELLITUS
; TITLE OF INVENTION: BY ADMINISTERING A PSEUDO-ISLET LIKE AGGREGATE DIFFERENTIATED FROM
; TITLE OF INVENTION: A NESTIN-POSITIVE PANCREATIC STEM CELL
; FILE REFERENCE: 3284/1223

```

; CURRENT APPLICATION NUMBER: US/10/984,645
; CURRENT FILING DATE: 2004-11-09
; PRIOR APPLICATION NUMBER: US 09/731,255
; PRIOR FILING DATE: 2000-12-06
; PRIOR APPLICATION NUMBER: US 60/169,082
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: US 60/215,109
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: US 60/239,880
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 1618
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-984-645-2

```

Query Match	5.1%;	Score 117;	DB 1;	Length 1618;
Best Local Similarity	24.2%;	Pred. NO. 0.5;		
Matches 75; Conservative	37;	Mismatches 110;	Indels 88;	Gaps 15;

```

Qy 4 QOSSPLVKFASPAESFUSAGPDNFTSL--FADSTPTNLNPRDMMTPOSDAIDSRSLSVI 61
Db 694 ENQBFLSRLEDENKEAFRSLEKENQEPLKLTLEEDQDSIVRPLETENHKSLSRLS----- 746
Qy 62 PESQQAEDSHSHTSATAPSTSEKKPKVKRKSQGQ-----VLPEPKTNLPPrKRAKTEDE 116
Db 747 -EEQDQE-----TLRLTEKETQORRSLSGEQDQWTLRPPEKVDLEP---LKSLDQ 792
Qy 117 KEQRRVVRVLNRRRAOSSRR-----KRLEVEALEK-----RNKLETLININVQTNLIL 167
Db 793 EIAKPLN--ENQBFLSKLKEESVEAVKLSLETEILESLSKSAQENLETKSPETQAPLWT 850
Qy 168 VEELNRRRRSGVVTRSSSLDLSLQDSITLSCQLFGSRDQGTMSNPEQSLMQIMRSAAN 227
Db 851 PEEIN--KSGG--NESSRKGNSTTGVCGSE---PRDIQTPGRGESGIIIE----- 893
Qy 228 PTVNPASLSPLPP-----IS---DKPEQTKEEDREQADEEME----- 264
Db 894 -----ISGMEPGPEFISRGVDKESQORNLEEBENLKGYESLRLSLEEGQELPQSA 946
Qy 265 --QTWHETKE 272
Db 947 DVORWEDTKE 956

```

```

RESULT 11
US-10-821-234-963
; Sequence 963, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Presclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 963
; LENGTH: 703
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-963

```

Query Match 4.9%; Score 113.5; DB 1; Length 703;
Best Local Similarity 18.9%; Pred. No. 0.31;
Matches 72; Conservative 62; Mismatches 146; Indels 101; Gaps 14;

```
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 16
; LENGTH: 691
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-131-826A-16

Query Match      4.9%; Score 111.5; DB 1; Length 691;
Best Local Similarity 19.0%; Pred. No. 0.42;
Matches 72; Conservative 61; Mismatches 145; Indels 101; Gaps 14;

QY 13 EASPAEFLSAPGDNFTSLFADSTPST-LNPRDMTPDSVADIDSRLSVIP-ESQDAEDD 70
Db 2 EESPLSRAPSRGGVNFNLNARTYIPNTKVECHYTLPGTMPASADWIGIFKVEACVRDY 61
QY 71 ESHSTSATAPSTSEKKPVKKRSW-GQVLPPEPKTNL-----PPKRAKT 113
Db 62 HTFWSSVPSTTGGSPHTSVQFQASVLPKPGAQLYQFVYVNRQGVCGQSPPPQPREP 121
QY 114 EDEKEQRRVER-----VLNRRAAQSSRRKRL-----EVEALEKR 149
Db 122 RPDMLVTLERADGGSDILLVVPKATVLQN-QLDESOQERNDLMQLKQLEGQVTELSR 180
QY 150 NKELETLLINVOKTNLLIIVELNFRSSGVVTRSSPLD-----SLQDSI-TL 197
Db 181 VOELERALARATAREHTELMEOYKIGISRSHGITEERDILSRQOGDHVARILEDDIOTI 240
QY 198 SQQLFGS-----RDG-QTMSNPEQSLMDQIMRSAAANTVYNPASLSPLPISKEFQT 249
Db 241 SEKVLTKVELDLRLDTVKALTREQKLLGLKEVQADKEQ-----SEALEQV 288
QY 250 KEEDEEQADEDEMEQTHETKEAAAAKEKN-----SKQSRVS----- 287
Db 289 AQENHHLNLDLKEAKSWOEQSAQAQRLKDKVAQMDTLGQAQQRVAEPLKEQLRGA 348
QY 288 ----TDSTORPAVSIGGDA 303
Db 349 QELAASSQOKATLIGELA 367

RESULT 13
US-10-821-234-1477
; Sequence 1477, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; PRIOR FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt seq_genes Version 1.0
; SEQ ID NO 1477
; LENGTH: 667
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1477

Query Match      4.8%; Score 111; DB 1; Length 667;
Best Local Similarity 21.0%; Pred. No. 0.43;
Matches 101; Conservative 57; Mismatches 168; Indels 154; Gaps 21;

QY 5 QSSPLVKFEAPSAEFLSAPGDNFTSLFADSTPSTLNPRDMTPDSVADIDSRLSVIPES 64
```

```
11 KFEASPAEFLSAPGDNFTSLFADSTPST-LNPRDMTPDSVADIDSRLSVIP-ESQDAE 68
Db 12 RMESPLSRAPSRGGVNFNLNARTYIPNTKVECHYTLPGTMPASADWIGIFKVEACVR 71
QY 69 DDESHSTSATAPSTSEKKPVKKRSW-GQVLPPEPKTNL-----PPKRA 111
Db 72 DYHTFWSSVPSTTGGSPHTSVQFQASVLPKPGAQLYQFVYVNRQGVCGQSPPPQPRE 131
QY 112 KTEDEKEQRRVER-----VLNRRAAQSSRRKRL-----EVEALE 147
Db 132 EPRPDMLVTLERADGGSDILLVVPKATVLQN-QLDESOQERNDLMQLKQLEGQVTELR 190
QY 148 KRNELETLLINVOKTNLLIIVELNFRSSGVVTRSSPLD-----SLQDSI- 195
Db 191 SRVOELERALARATAREHTELMEOYKIGISRSHGITEERDILSRQOGDHVARILEDDIQ 250
QY 196 TLSQQLFGS-----RDG-QTMSNPEQSLMDQIMRSAAANTVYNPASLSPLPISKEF 247
Db 251 TISEKVLTKVELDLRLDTVKALTREQKLLGLKEVQADKEQ-----SEALE 298
QY 248 QTKEDEEQADEDEMEQTHETKEAAAAKEKN-----SKQSRVS----- 287
Db 299 QVAQENHHLNLDLKEAKSWOEQSAQAQRLKDKVAQMDTLGQAQQRVAEPLKEQLR 358
QY 288 ----TDSTORPAVSIGGDA 303
Db 359 QELAASSQOKATLIGELA 379

RESULT 12
US-10-131-826A-16
; Sequence 16, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
```

```
Db 222 QGACTVRSAPQOEKKRKVEKMTMSQDALEALSASLGTRQ-----APFLDLSRIKEV 274
QY 65 QDA-----EDESHSTSATAPSTSEKPKVKRKSWGQVLPEPKTNLPKRRAK- 112
Db 275 DEAKAKEEKLEKGEDDE-----TIPSEYRLKPAITD-KDGKPLLPPEPEKPKRSESEL 327
QY 113 ----TED-----EKEQRRVERVLNRRRAAQS--SRERKRRLVEALEKRNKELETLIN 159
Db 328 IDLSDFDRSECKEKPSTKTEESKAAAPAPVSEAVCRTSMCSIQAPPEPATL--- 384
QY 160 VQKTNLILVEELNRRFRSSGVVTRSSPLDSIQDSTLTSQQLFGSRDGTMSNPQSLMD 219
Db 385 -----KGTIV--PDDAVEALADSL-----GKEADPEDG--- 410
QY 220 QIMRSAANTVNPASLSPSLPISDK-EFQTKEEDEEQADEDEMEQWTHETKEAAAAKE 278
Db 411 -----KPVMDKVKKEKEDEKLEKGEETIPPDYRLKEE---VKD 447
QY 279 KNSKQSRVSTSTORPAVSIIGG--DAAVPVFSDDAGANCLGLDPVHQDDGPFSGHSGFL 336
Db 448 KDGKPLLPKESKEQFPFMSDFELDLALSDFGSQPNASSLKFEADK-----L 494
QY 337 SAALDADRYLLESQLLASPNASTVDDDYLAGDSAACTFNPLPSDYDF-DINDFLTD---- 391
Db 495 AAAI-----SEVVSQTPASTTQAGAPRPTS-----QSDKDLDDLKLSLQ 539
QY 392 -----DANHAAYDIVAASNAADRELDEIHDPENQIPSRHSIQOQSGASHGCDGG 446
Db 540 RQDPDPENKPMEDKVKKEK-----AKAHRDLKERDDTIPPEY-----RHLLDDNG 585

RESULT 14
US-10-821-234-1514
; Sequence 1514, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; PRIOR FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes version 1.0
; SEQ ID NO 1514
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1514

Query Match 4.8%; Score 111; DB 1; Length 915;
Best Local Similarity 22.6%; Pred. No. 0.63;
Matches 82; Conservative 52; Mismatches 145; Indels 84; Gaps 15;

QY 29 TSLPADSTPSTLN-PRDMWTPDSVADIDSRLSVIPESQ-----DABDDSHSTSATAP 80
Db 501 TSSGKSTITRVNSPGLTARLGSVTHVTSFSHAPFSRGGCSIKWEAPPAEPLAAVRA 560
QY 81 STSEKKPKVKRKSQGVLPPEKTNLPKRRAKTEDEK-----EQRVERV--- 125
Db 561 NGAQTRVYK-----APEGRSPLSABELMTIEDEGLDKMLDQSTDFEERKLIRAA 613
QY 126 -LRNPRRAQSRERKRLVEALEKRNKELETLINQVQTNLILVEELNRRFRSSGVVTR 184
Db 614 ELRQKRDRQDKERERRRLEQEA---RGREFE-----GRGNTATETTTTR 653
QY 185 SSPL--DSLQDSITLSQQLFGSRDGTMSNPQSLMDQIMRSAAN-----PTVNP 233
```

Search completed: November 23, 2005, 03:32:37
Job time : 7.71544 secs

```
Db 654 SORAAADGSAVSTVTYKTERLHVHNDG-TRTARTTTVESSFVRSENGSGSTMOTKTFSSS 712
QY 234 SLSPSLPISDKPEFQTKEE-----DEEQADEDEM--EQTWTHETKEAAAAK---EKNS 281
Db 713 SSSKMGSIFFREDQASPRAGSLAALEKRAQAEKKELMKAQSLPKTTSASQARKAMIEKLE 772
QY 282 KQSRVSTSTORPAVSIIGGDAVP-----VPSDDAGANCLGLDPVHQDDGPFSGHSGF 334
Db 773 KEGAAGSGGPRAAVQVSTSGFVNPANSIKOMLLDWCRAKTRGYE--HVDIQNFSSSWSD 830
QY 335 GLS 337
Db 831 GMA 833

RESULT 15
US-10-793-626-1780
; Sequence 1780, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1780
; LENGTH: 1155
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1780

Query Match 4.8%; Score 110; DB 1; Length 1155;
Best Local Similarity 23.2%; Pred. No. 0.98;
Matches 92; Conservative 43; Mismatches 190; Indels 72; Gaps 14;

QY 63 ESQDAEDDESHTSATAPSTSEKPKVKRKSQW---GQVLPEKTNLPKRRAK----- 112
Db 49 EAKAEDNQLESASKEEQKGRDNESSKLNQVLDNGSHSEKTTNVNNAETVKVKEAPT 108
QY 113 TEDEKQRRVERVLNRRRAAQSRRERKRLVEALEKRNKELETLINQVQTNLILVEELN 172
Db 109 TSDVSKPKANEAVTNESTKPKTTEAPTVEESIAETPKTSTTQODSTEKNPSLKDNLN 168
QY 173 RFRSSGVVTRSSPLDSLQDSITLSQQLFGSRDGTMSNPQSLMDQIMRSAAN--PTV 230
Db 169 -----SSSTTSKESKTD--EHSITKQAQMSNTKNSLNDTNDSPQTS---EKTSSQANNDSTD 218
QY 231 NPASLSPL--PPISDKPEFQTKEEDEEQADEDEMEQWTHETKEAAAAKKNKQSRVST 288
Db 219 NQSAQSKQLDSKPSQKQVYKTKFNDE--PTQDQVHTHTKLKTPSISTDSSVNDKQ----- 271
QY 289 DSTORPAVSIIGGDAVPVPSDDAGANCLGL-----DPVHQDDG 326
Db 272 DYTSASVASLGVDSNETEATINAVRDNLDLKAASREQINEAIIAEALKKDSNPDPYDVT 331
QY 327 PFTSGHSGFL-SAALDADRYLLESQLLASPNASTVDDDYLAGDSAACTFNLPSPDYDFDI 385
Db 332 PLALNTSQSKNSPHKSASPRMNLMSLAAEPNSGKNVNDKVK-----ITNPTLS----- 379
QY 386 NDFITDDANHAAYDIVAASN---YAAADRELDEIHD 419
Db 380 ---LNKSNHANNHWIPTSNEQFNLKANYELDDSDSIKE 413
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: November 23, 2005, 03:12:56 ; Search time 146.003 Seconds
(without alignments)
1290.661 Million cell updates/sec

Title: US-10-663-450-2

Perfect score: 2294

Sequence: 1 MAFQSSPLVKFEASPAESF.....QPQSGASSHCDDGGIAGV 451

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 41829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.Main:

- 1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pap:*
- 2: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pap:*
- 3: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pap:*
- 4: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pap:*
- 5: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pap:*
- 6: /cgn2_6/prodata/1/pubpaa/US11_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2294	100.0	451	3	US-09-816-277-2
2	2294	100.0	451	3	US-09-816-277-5
3	2294	100.0	451	4	US-10-663-450-2
4	2294	100.0	451	4	US-10-663-450-5
5	434.5	18.9	386	3	US-09-816-277-16
6	434.5	18.9	386	4	US-10-663-450-16
7	434	18.9	349	3	US-09-816-277-4
8	434	18.9	349	3	US-09-816-277-6
9	434	18.9	349	4	US-10-663-450-4
10	434	18.9	349	4	US-10-663-450-6
11	429.5	18.7	409	3	US-09-533-029-104
12	429.5	18.7	409	3	US-10-295-403-158
13	429.5	18.7	409	4	US-10-412-6998-684
14	426.5	18.6	342	3	US-09-816-277-19
15	426.5	18.6	342	4	US-10-663-450-19
16	312.5	13.6	174	4	US-10-767-701-61238
17	312.5	10.1	200	4	US-10-369-493-21867
18	196	8.5	68	3	US-09-816-277-60
19	196	8.5	68	4	US-10-663-450-60
20	166	7.2	3257	6	US-11-097-143-29298
21	165.5	7.2	192	3	US-09-772-656-6
22	165.5	7.2	192	3	US-09-772-656-10
23	165.5	7.2	192	4	US-10-396-199A-6
24	165.5	7.2	192	4	US-10-396-199A-10
25	163.5	7.1	1442	5	US-10-450-763-55960
26	163.5	7.1	2408	5	US-10-450-763-40990
27	163	7.1	192	3	US-09-772-656-2

28	163	7.1	192	4	US-10-396-199A-2	Sequence 2, Appli
29	148.5	6.5	650	3	US-09-801-368-430	Sequence 430, App
30	148.5	6.5	2048	3	US-09-815-379-12	Sequence 12, Appl
31	148.5	6.5	2057	3	US-09-815-379-10	Sequence 10, Appl
32	147.5	6.4	2058	3	US-09-815-379-17	Sequence 17, Appl
33	147.5	6.4	2058	4	US-10-021-660-120	Sequence 120, App
34	147.5	6.4	2058	4	US-10-211-462-207	Sequence 207, App
35	147.5	6.4	2058	4	US-10-648-593-188	Sequence 188, App
36	147.5	6.4	2058	5	US-10-723-860-742	Sequence 742, App
37	146.5	6.4	1017	4	US-10-032-585-7590	Sequence 7590, Ap
38	145.5	6.3	639	5	US-10-697-787-14	Sequence 14, Appl
39	145.5	6.3	639	5	US-10-512-600-14	Sequence 14108, A
40	145.5	6.3	5533	6	US-11-097-143-24108	Sequence 24108, A
41	145.5	6.3	5560	4	US-10-263-929-142	Sequence 142, App
42	145.5	6.3	5560	6	US-11-097-143-40272	Sequence 40272, A
43	145	6.3	1000	4	US-10-128-714-3305	Sequence 3305, Ap
44	144.5	6.3	665	4	US-10-425-115-190929	Sequence 190929,
45	144.5	6.3	672	4	US-10-425-114-58469	Sequence 58469, A

ALIGNMENTS

RESULT 1

US-09-816-277-2

; Sequence 2, Application US/09816277

; Publication No. US20010034045A1

; GENERAL INFORMATION:

; APPLICANT: Penttila, Merja E.

; APPLICANT: Ward, Michael

; APPLICANT: Wang, Huaming

; APPLICANT: Valkonen, Mari J.

; APPLICANT: Saloheimo, Markku

; TITLE OF INVENTION: Increased Production of Secreted

; FILE OF INVENTION: Proteins by Recombinant Eukaryotic Cells

; FILE REFERENCE: GC590-2

; CURRENT APPLICATION NUMBER: US/09/816,277

; PRIOR FILING DATE: 2001-03-23

; PRIOR APPLICATION NUMBER: US 09/534,692

; PRIOR FILING DATE: 2000-03-24

; NUMBER OF SEQ ID NOS: 63

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 451

; TYPE: PRT

; ORGANISM: Trichoderma reesei

; US-09-816-277-2

Query Match 100.0%; Score 2294; DB 3; Length 451;
Best Local Similarity 100.0%; Pred. No. 2.6e-143; Indels 0; Gaps 0;
Matches 451; Conservative 0; Mismatches 0;

QY	1	MAFQSSPLVKFEASPAESFSLGAPGNFTSLFADSTPSTLNPRDMMTPDSVADIDRLSV	60
DB	1	MAFQSSPLVKFEASPAESFSLGAPGNFTSLFADSTPSTLNPRDMMTPDSVADIDRLSV	60
QY	61	IPESQDAEDDESHSTSATPSTSEKPKVKRKSQGVLPKPKTNLPKRKAETDEKEQR	120
DB	61	IPESQDAEDDESHSTSATPSTSEKPKVKRKSQGVLPKPKTNLPKRKAETDEKEQR	120
QY	121	RVERVLNRRRAQSSRRERKLEVEALEKKNKELETLLINVQKTNLLIIVELNFRSSGV	180
DB	121	RVERVLNRRRAQSSRRERKLEVEALEKKNKELETLLINVQKTNLLIIVELNFRSSGV	180
QY	181	VTRSSPLDLSODITLSQOLFGRDQGTWNSPEQSLMDQIMRSAANPTVNPASLPSLP	240
DB	181	VTRSSPLDLSODITLSQOLFGRDQGTWNSPEQSLMDQIMRSAANPTVNPASLPSLP	240
QY	241	PISDKFQTKEDDEQADEEMEQTWHTKEAAAAKKNKSKOSRVSTSTQTPAVSIGG	300
DB	241	PISDKFQTKEDDEQADEEMEQTWHTKEAAAAKKNKSKOSRVSTSTQTPAVSIGG	300
QY	301	DRAVPFSDAGANCLGLDPVHQDDPFSGHSGFGLSAAALDADRYLLESQALLASPNASTV	360

Db 301 DAAVPVFSDDAGANCLGLDPVHQDDGPFSGHSGFGLSAALDADRYLLESQLLASPNASTV 360
QY 361 DDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAADRELDLEIHP 420
Db 361 DDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAADRELDLEIHP 420
QY 421 ENQIPSRHSIIQQPSGASHGCGDDGGIAGV 451
Db 421 ENQIPSRHSIIQQPSGASHGCGDDGGIAGV 451

RESULT 2

US-09-816-277-5
; Sequence 5, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-09-816-277-5

Query Match 100.0%; Score 2294; DB 3; Length 451;
Best Local Similarity 100.0%; Pred. No. 2.6e-143;
Matches 451; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAFQSSPLVKFEASPAESFLSAPGDNFTSLPADSTPSTLNPRDMTPPSVADIDSRLSV 60
Db 1 MAFQSSPLVKFEASPAESFLSAPGDNFTSLPADSTPSTLNPRDMTPPSVADIDSRLSV 60
QY 61 IPESQDAEDDESHSTSATAPSTSEKKPVKKRSGQVLPPEKTNLPKRKTEDEKEOR 120
Db 61 IPESQDAEDDESHSTSATAPSTSEKKPVKKRSGQVLPPEKTNLPKRKTEDEKEOR 120
QY 121 RVERVLNRRAAQSSRKRLEVEALEKRNKELETLLINVQKTNLILVEELNFRFRSSGV 180
Db 121 RVERVLNRRAAQSSRKRLEVEALEKRNKELETLLINVQKTNLILVEELNFRFRSSGV 180
QY 181 VTRSSPLDSLQDSITLSQQLFGSRDGTMSNPEQSLMDQIMRSAANPTVNPASLPSLP 240
Db 181 VTRSSPLDSLQDSITLSQQLFGSRDGTMSNPEQSLMDQIMRSAANPTVNPASLPSLP 240
QY 241 PISDKFQTKERDEEQADEMEQTHETKEAAAKENKSKQSRVSTDTSTORPAVSIG 300
Db 241 PISDKFQTKERDEEQADEMEQTHETKEAAAKENKSKQSRVSTDTSTORPAVSIG 300
QY 301 DAAVPVFSDDAGANCLGLDPVHQDDGPFSGHSGFGLSAALDADRYLLESQLLASPNASTV 360
Db 301 DAAVPVFSDDAGANCLGLDPVHQDDGPFSGHSGFGLSAALDADRYLLESQLLASPNASTV 360
QY 361 DDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAADRELDLEIHP 420
Db 361 DDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAADRELDLEIHP 420
QY 421 ENQIPSRHSIIQQPSGASHGCGDDGGIAGV 451
Db 421 ENQIPSRHSIIQQPSGASHGCGDDGGIAGV 451

RESULT 3

US-10-663-450-2
; Sequence 2, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-663-450-2

Query Match 100.0%; Score 2294; DB 4; Length 451;
Best Local Similarity 100.0%; Pred. No. 2.6e-143;
Matches 451; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAFQSSPLVKFEASPAESFLSAPGDNFTSLPADSTPSTLNPRDMTPPSVADIDSRLSV 60
Db 1 MAFQSSPLVKFEASPAESFLSAPGDNFTSLPADSTPSTLNPRDMTPPSVADIDSRLSV 60
QY 61 IPESQDAEDDESHSTSATAPSTSEKKPVKKRSGQVLPPEKTNLPKRKTEDEKEOR 120
Db 61 IPESQDAEDDESHSTSATAPSTSEKKPVKKRSGQVLPPEKTNLPKRKTEDEKEOR 120
QY 121 RVERVLNRRAAQSSRKRLEVEALEKRNKELETLLINVQKTNLILVEELNFRFRSSGV 180
Db 121 RVERVLNRRAAQSSRKRLEVEALEKRNKELETLLINVQKTNLILVEELNFRFRSSGV 180
QY 181 VTRSSPLDSLQDSITLSQQLFGSRDGTMSNPEQSLMDQIMRSAANPTVNPASLPSLP 240
Db 181 VTRSSPLDSLQDSITLSQQLFGSRDGTMSNPEQSLMDQIMRSAANPTVNPASLPSLP 240
QY 241 PISDKFQTKERDEEQADEMEQTHETKEAAAKENKSKQSRVSTDTSTORPAVSIG 300
Db 241 PISDKFQTKERDEEQADEMEQTHETKEAAAKENKSKQSRVSTDTSTORPAVSIG 300
QY 301 DAAVPVFSDDAGANCLGLDPVHQDDGPFSGHSGFGLSAALDADRYLLESQLLASPNASTV 360
Db 301 DAAVPVFSDDAGANCLGLDPVHQDDGPFSGHSGFGLSAALDADRYLLESQLLASPNASTV 360
QY 361 DDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAADRELDLEIHP 420
Db 361 DDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAADRELDLEIHP 420
QY 421 ENQIPSRHSIIQQPSGASHGCGDDGGIAGV 451
Db 421 ENQIPSRHSIIQQPSGASHGCGDDGGIAGV 451

RESULT 4

US-10-663-450-5
; Sequence 5, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku


```

; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-663-450-5

Query Match          100.0%; Score 2294; DB 4; Length 451;
Best Local Similarity 100.0%; Pred. No. 2,6e-143;
Matches 451; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAFOQSPLVKFEASPAESFLSAPGDNFTSLFADSTPSTLNPRDMWTPDSVADIDSRSLV 60
Db 1 MAFOQSPLVKFEASPAESFLSAPGDNFTSLFADSTPSTLNPRDMWTPDSVADIDSRSLV 60

Qy 61 IPESQDAEDDESHSTSATAPSTSEKKPVKKKSKGWQVLPEPKTNLPPRKRAKTEDEKEQR 120
Db 61 IPESQDAEDDESHSTSATAPSTSEKKPVKKKSKGWQVLPEPKTNLPPRKRAKTEDEKEQR 120

Qy 121 RVERVLNRRRAAQSSRRKRLEVEALEKRNKELETLINVKTNLILVEELNRRFRSSGV 180
Db 121 RVERVLNRRRAAQSSRRKRLEVEALEKRNKELETLINVKTNLILVEELNRRFRSSGV 180

Qy 181 VTRSSSLDLSQDSITILSQQLFGSRDGTMSNPNPQSLMDQIMRSAANPTVNPASLSPLP 240
Db 181 VTRSSSLDLSQDSITILSQQLFGSRDGTMSNPNPQSLMDQIMRSAANPTVNPASLSPLP 240

Qy 241 PISDKFQTKXEDEBEQADEBEEMQTHWETKEAAAAKSKSKSRVSTDSQRPASVSTGG 300
Db 241 PISDKFQTKXEDEBEQADEBEEMQTHWETKEAAAAKSKSKSRVSTDSQRPASVSTGG 300

Qy 301 DAAVPVFSDDAGANCLGLDPVHQDDGPFSGHSGFLSNAALDADRYLLESQLLASPNASTV 360
Db 301 DAAVPVFSDDAGANCLGLDPVHQDDGPFSGHSGFLSNAALDADRYLLESQLLASPNASTV 360

Qy 361 DDDYLAGDSAAACFTNPLPSDYDFDINDFLTDDANHAAYDIVAAASNYAAAARELDLEIHP 420
Db 361 DDDYLAGDSAAACFTNPLPSDYDFDINDFLTDDANHAAYDIVAAASNYAAAARELDLEIHP 420

Qy 421 ENQIPSRHSIQQPQSGASSHGCDGGIAVG 451
Db 1
Db 421 ENQIPSRHSIQQPQSGASSHGCDGGIAVG 451

RESULT 5
US-09-816-277-16
; Sequence 16, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 386

```

```

; TYPE: PRT
; ORGANISM: Aspergillus niger
US-09-816-277-16

Query Match      18.9%; Score 434.5; DB 3; Length 386;
Best Local Similarity 32.3%; Pred. No. 1.4e-20;
Matches 147; Conservative 50; Mismatches 139; Indels 119; Gaps 20;

Qy 24 PGDNFTSLFADSTP-STLNPRDMWTDPDSVADIDS-----RLSVIPESQDAEDDES 72
Db 22 PGITWPGCLAPPSPVMTRWPFLMMEAFSPVDLSAGSPTPELPLLTVSADTSLDSD-- 79
Qy 73 HSTSATAPSTSEKKPKVKRKSGWGVLPEPKTNLPPrKRAKTEDEKQRRVERVLNRRAA 132
Db 80 -SVQAGETKAEEKPKVKRKSGWGQLPVPKTNLPPrKRAKTEDEKEQRIRVLRNAAA 138
Qy 133 OSSRERKRLEVEALEKENKELETLLINVQKTNLILVEELNFRSSGVVTRSSPLDSLQ 192
Db 139 QTSRERKRLEWEKLENEK-----IQMEQQOQFLQLRSQMAENRNLM-----Q 182
Qy 193 DSITTLSQLFGSRDGQTMSNPQEOSLMDQIMRSAANPTVNPAISPSLPPISDKFEQTKEE 252
Db 183 QVAQLSAEVRGSR-GNT---PKPG-----SPVASPTLPTLFKQ-----ER 220
Qy 253 DE---BOADEBEMEQTHETKEAAAKENKSQSVSTDTSTORPAVISGG-----DAAVP 305
Db 221 DEIPLERIPTPTSITDYSTPLRPSTLAEE-----SSDVTQHPAVSVAGLEGEGSALS 272
Qy 306 VFSDGAGACGLHDHPVHDDGPFSIGHSGLSAALDADRY--LLESQLLASPNASTVDD- 362
Db 273 LF-DVGSN---PEPHAADD-----LAAPLSDDDFHRLFNVNDSPVGSOSSVLEDG 317
Qy 363 ---DYLAGDSAACFTNPLPSDYDINDFLTDDHAAYDIVAASNVAADRELDLEIHD 419
Db 318 FAFDVLDGDLFAF--PFDSWVDF-----PESVPEGIE 350
Qy 420 PENQIFSRHSIQ----QPQSASSHGCDGDIAGV 450
Db 351 PPGLPDTSRQTSVVQPSLGSRCSRCDGGIAAG 385

RESULT 6
US-10-663-450-16
; Sequence 16, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCES: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-10-663-450-16

Query Match      18.9%; Score 434.5; DB 4; Length 386;
Best Local Similarity 32.3%; Pred. No. 1.4e-20;
Matches 147; Conservative 50; Mismatches 139; Indels 119; Gaps 20;

Qy 24 PGDNFTSLFADSTP-STLNPRDMWTDPDSVADIDS-----RLSVIPESQDAEDDES 72
Db 22 PGITWPGCLAPPSPVMTRWPFLMMEAFSPVDLSAGSPTPELPLLTVSADTSLDSD-- 79

```

```
QY 73 HSTSATAPSTSEKKPVKKRKSQGVLPKPTNLPPKRAKTEDEKEQRRVERVLNRRAA 132
Db 80 -SVQAGETKAEKKPVKKRKSQGVLPKPTNLPPKRAKTEDEKEQRIERVLRRAA 138
QY 133 QSRERKRLEVEALEKRNKELETLINVKNTLILVEELNRRFRSSGVTVTRSSSPDLSIQ 192
Db 139 QTSRERKRLEMEKLENEK-----IQMEQONQFLQLRQSMAEENRLN-----Q 182
QY 193 DSITLSQQLFGSRDGTMSNPEQSLMDQIMRSAANTVNPASLSPSLPTSDKEFQTKEE 252
Db 183 QVAQLSAEVRGSR-GNT---PKRG-----SPVSAASPTLTPTLFKQ-----ER 220
QY 253 DE---BOADEDEMEQTHETKEAAAKKNSQSRVSTDSTQRPVAVSIGG-----DAAVP 305
Db 221 DEIPLERIPPTPSITDYSPTLRPSLAE-----SSDVTQHPAVSVAGLEGESALS 272
QY 306 VFSDDAGANCLGLDPVHQDDGPPSIGHSGLSAALDADRY--LLESQLLASPNASTVDD- 362
Db 273 LF--DVGSN---PEPHAADD-----LAAPLSDDDFHRLFNVDSPVGSDDSVLEDG 317
QY 363 ---DYLAGDSAACTNPLPSDYPDINDFLTDDANHAAYDIVAASNAAADRELDLIHD 419
Db 318 FAFVDLVDGDLGSAF--PFDGMVDFD-----PESVGFEGIE 350
QY 420 PENQIPSRHSIQ---QPSQGASSHGCDGDIAGV 450
Db 351 PPHGLPDETSTSRQTSVQPSLGASTSRCDGGGIAAG 385

RESULT 7
US-09-816-277-4
; Sequence 4, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-816-277-4

Query Match 18.9%; Score 434; DB 3; Length 349;
Best Local Similarity 32.4%; Pred. No. 1.3e-20;
Matches 149; Conservative 50; Mismatches 109; Indels 152; Gaps 21;

QY 26 DNFTSL-----PADS--TPS-----TLNPRD--MMTPDSVADIDSRLSVIPESQDAEDD 70
Db 5 DRFSPVKMEDAFANSPPTPSLEVPVLTSPADTSLRTKNVQA-----47
QY 71 ESHSTSATAPSTSEKKPVKKRKSQGVLPKPTNLPPKRAKTEDEKEORRVERVLNR 130
Db 48 -----TKP--EEKKPAKKRKSQGVLPKPTNLPPKRAKTEDEKEORRVERVLNR 98
QY 131 AAQSRERKRLEVEALEKRNKELETLINVKNTLILVEELNRRFRSSGVTVTRSSPLDS 190
Db 99 AAQTSRERKRLEMEKLESEK-----IDMEQONQFLQLRQAQMEANNRLS-----143
QY 191 LQDSITLSQQLFGSRDGTMSNPEQSLMDQIMRSAANTVNPASLSPSLPTSDKEFQTK 250
Db 144 -QVAQLSAEVRGSRH-----STPTSS-----SPASVSPTLTPTLFKQ-----180
QY 251 EDEDEQADE---DEMEQTHETKEAAAKKNSQSRVSTDSTQRPVAVSIGG-----DAA 303
Db 181 EGDEVPLDRIPPTPSITDYSPTLRPSLAE-----SPDLTQHPAVSVGLEGDESA 232
QY 304 VPVFSDDAGANCLGLDPVHQDDGPPS---IGHSGLSAALDADRYLLE-----SQ 350
Db 233 LTLF--DLGAS--IKHEPTHDLTAPLSDDDFRFLFNGDSSLESDDLSDGFAFDVLDSGD 289
QY 351 LLASPNASTVDDYLAGDSAACTNPLPSDYPDINDFLTDDANHAAYDIVAASNAAAD 410
```

```
Db 144 -QVAQLSAEVRGSRH-----STPTSS-----SPASVSPTLTPTLFKQ-----180
QY 251 EDEDEQADE---DEMEQTHETKEAAAKKNSQSRVSTDSTQRPVAVSIGG-----DAA 303
Db 181 EGDEVPLDRIPPTPSITDYSPTLRPSLAE-----SPDLTQHPAVSVGLEGDESA 232
QY 304 VPVFSDDAGANCLGLDPVHQDDGPPS---IGHSGLSAALDADRYLLE-----SQ 350
Db 233 LTLF--DLGAS--IKHEPTHDLTAPLSDDDFRFLFNGDSSLESDDLSDGFAFDVLDSGD 289
QY 351 LLASPNASTVDDYLAGDSAACTNPLPSDYPDINDFLTDDANHAAYDIVAASNAAAD 410

RESULT 8
US-09-816-277-6
; Sequence 6, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-816-277-6

Query Match 18.9%; Score 434; DB 3; Length 349;
Best Local Similarity 32.4%; Pred. No. 1.3e-20;
Matches 149; Conservative 50; Mismatches 109; Indels 152; Gaps 21;

QY 26 DNFTSL-----PADS--TPS-----TLNPRD--MMTPDSVADIDSRLSVIPESQDAEDD 70
Db 5 DRFSPVKMEDAFANSPPTPSLEVPVLTSPADTSLRTKNVQA-----47
QY 71 ESHSTSATAPSTSEKKPVKKRKSQGVLPKPTNLPPKRAKTEDEKEORRVERVLNR 130
Db 48 -----TKP--EEKKPAKKRKSQGVLPKPTNLPPKRAKTEDEKEORRVERVLNR 98
QY 131 AAQSRERKRLEVEALEKRNKELETLINVKNTLILVEELNRRFRSSGVTVTRSSPLDS 190
Db 99 AAQTSRERKRLEMEKLESEK-----IDMEQONQFLQLRQAQMEANNRLS-----143
QY 191 LQDSITLSQQLFGSRDGTMSNPEQSLMDQIMRSAANTVNPASLSPSLPTSDKEFQTK 250
Db 144 -QVAQLSAEVRGSRH-----STPTSS-----SPASVSPTLTPTLFKQ-----180
QY 251 EDEDEQADE---DEMEQTHETKEAAAKKNSQSRVSTDSTQRPVAVSIGG-----DAA 303
Db 181 EGDEVPLDRIPPTPSITDYSPTLRPSLAE-----SPDLTQHPAVSVGLEGDESA 232
QY 304 VPVFSDDAGANCLGLDPVHQDDGPPS---IGHSGLSAALDADRYLLE-----SQ 350
Db 233 LTLF--DLGAS--IKHEPTHDLTAPLSDDDFRFLFNGDSSLESDDLSDGFAFDVLDSGD 289
QY 351 LLASPNASTVDDYLAGDSAACTNPLPSDYPDINDFLTDDANHAAYDIVAASNAAAD 410
```

```
Db 290 LSAPFPDSVDFD-----TEPVTLE-DLEQTNGLSDSASCKAASL----- 328
Qy 411 RELDLIHPDNPQIPSRHSIQPQSGASSHGCDGGIAGV 450
Db 329 -----QPSHGASTSRCDDGGIAG 347

RESULT 9
US-10-663-450-4
; Sequence 4, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GCS90-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; PRIOR FILING DATE: 2003-09-15
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-663-450-4

Query Match 18.9%; Score 434; DB 4; Length 349;
Best Local Similarity 32.4%; Pred. No. 1.3e-20;
Matches 149; Conservative 50; Mismatches 109; Indels 152; Gaps 21;

Qy 26 DNFTSL-----PADS-----TPS-----TLNPRD--MWTSPSVADIDSRLSVIPESQDAEDD 70
Db 5 DRFSPVKMEDAFANSPPTPSLEVPVLTVPADTSLRTKNVVAQ----- 47
Qy 71 ESHSTSATAPSTSEKKPKVKRSGQVLPEPKTNLPKRKRAKTEDEKEORRVERVLNR 130
Db 48 -----TKP--EKKPAKPKRSGQELPVKTNLPKRKRAKTEDEKEORRVERVLNR 98
Qy 131 AAQSRERKRLEVEALEKRNKELETLINVKTNLILVEELNRRSSGVWTRSSPLDS 190
Db 99 AAQTSRERKRLEMEKLESEK-----IDMEQONQFLQLQAQMEANNRLS----- 143
Qy 191 LQDSITLSQQLFGSRDQGTMSNPESLMDQIMRSAAANTVNPASLSPSLPISDKKEFQTK 250
Db 144 -QQVAQLSAEVRGSRH-----STPTSS-----SPASVSPTLTPTLFKQ----- 180
Qy 251 EDEDEQADE--DEMEQTHETKEAAAANKSKQSRVSTSTORPAVSIIG--DAA 303
Db 181 EGDEVPLDRIPTPTSVTDYSFTLKPSLAE-----SPDLTQHPAVSVGGLEGDESA 232
Qy 304 VPFVSDDAGANCLGDPVHQDDGPFS---IGHSGLSAALDADRYLLE-----SQ 350
Db 233 LTLP--DLGAS-IXHEPTHDLTAPLSDDDFRFLFNGDSSLESDDLDDGFAFDVLDSDG 289
Qy 351 LLASPNASTVDDYLAGDSAACTNPLPSDYDFDINDFLTDANHAAYDIVAASNYAAAD 410
Db 290 LSAPFPDSVDFD-----TEPVTLE-DLEQTNGLSDSASCKAASL----- 328
Qy 411 RELDLIHPDNPQIPSRHSIQPQSGASSHGCDGGIAGV 450
Db 329 -----QPSHGASTSRCDDGGIAG 347
```

```
RESULT 10
US-10-663-450-6
; Sequence 6, Application US/10663450
```

```
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GCS90-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; PRIOR FILING DATE: 2003-09-15
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-663-450-6

Query Match 18.9%; Score 434; DB 4; Length 349;
Best Local Similarity 32.4%; Pred. No. 1.3e-20;
Matches 149; Conservative 50; Mismatches 109; Indels 152; Gaps 21;

Qy 26 DNFTSL-----PADS-----TPS-----TLNPRD--MWTSPSVADIDSRLSVIPESQDAEDD 70
Db 5 DRFSPVKMEDAFANSPPTPSLEVPVLTVPADTSLRTKNVVAQ----- 47
Qy 71 ESHSTSATAPSTSEKKPKVKRSGQVLPEPKTNLPKRKRAKTEDEKEORRVERVLNR 130
Db 48 -----TKP--EKKPAKPKRSGQELPVKTNLPKRKRAKTEDEKEORRVERVLNR 98
Qy 131 AAQSRERKRLEVEALEKRNKELETLINVKTNLILVEELNRRSSGVWTRSSPLDS 190
Db 99 AAQTSRERKRLEMEKLESEK-----IDMEQONQFLQLQAQMEANNRLS----- 143
Qy 191 LQDSITLSQQLFGSRDQGTMSNPESLMDQIMRSAAANTVNPASLSPSLPISDKKEFQTK 250
Db 144 -QQVAQLSAEVRGSRH-----STPTSS-----SPASVSPTLTPTLFKQ----- 180
Qy 251 EDEDEQADE--DEMEQTHETKEAAAANKSKQSRVSTSTORPAVSIIG--DAA 303
Db 181 EGDEVPLDRIPTPTSVTDYSFTLKPSLAE-----SPDLTQHPAVSVGGLEGDESA 232
Qy 304 VPFVSDDAGANCLGDPVHQDDGPFS---IGHSGLSAALDADRYLLE-----SQ 350
Db 233 LTLP--DLGAS-IXHEPTHDLTAPLSDDDFRFLFNGDSSLESDDLDDGFAFDVLDSDG 289
Qy 351 LLASPNASTVDDYLAGDSAACTNPLPSDYDFDINDFLTDANHAAYDIVAASNYAAAD 410
Db 290 LSAPFPDSVDFD-----TEPVTLE-DLEQTNGLSDSASCKAASL----- 328
Qy 411 RELDLIHPDNPQIPSRHSIQPQSGASSHGCDGGIAGV 450
Db 329 -----QPSHGASTSRCDDGGIAG 347

RESULT 11
US-09-533-029-104
; Sequence 104, Application US/09533029
; Publication No. US20030046723A1
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Broun, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Keddies, James
; APPLICANT: Pineda, Omaira
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-Liang
```

```

; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Reuber, Lynne
; TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES
; FILE REFERENCE: MBI-010
; CURRENT APPLICATION NUMBER: US/09/533,029
; CURRENT FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 60/125,814
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 104
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1034
US-09-533-029-104

Query Match      18.7%; Score 429.5; DB 3; Length 409;
Best Local Similarity 38.2%; Pred. No. 3.2e-20;
Matches 116; Conservative 39; Mismatches 70; Indels 79; Gaps 10;

QY 11 KFEASPA-----ESFLSAPGDNFTSLFADSTPTLNPRDMTPDVSADIDSLSVIPESQD 66
Db 8 KYENSPARTMVESFVSTPSSPHNPPLFD---NNLNPVDGFSF-----QS 48

QY 67 ABDESHSTATAPSTSEKKPVKKKSGQVLPEPKTNLPKRKRAKTEDEKEORRVERVL 126
Db 49 FDRDYNFNGSLGSLNLPKPKPIKKKSGQQLPEPKTNLPKRKRAKTODEKEORRVERVL 108

QY 127 RNRRAAQSSRRKRKEVEAL-----EKRKNELETLLINVOKTNLLILVEELNR---FRR 176
Db 109 RNRRAAQSSRRKRKEVEALEVEKRAIERKNWDLMEALADMEAKYLLQQLKCRASGYNK 168

QY 177 SSGVVTRSSSPDLSDLSQDS-----ITLSQQLFGSRDQGTWNSNPQSLMDQIMRSANPTVNP 232
Db 169 TNFLSYSDSTPDISDQSLPFTSKQLFNAQD--ELCRP---ISPOSIGPLTSRTVDP 223

QY 233 ASLSPSLPISDKFQTKDEEQADEMEQTHETKEAAAEEKSKQSRVSTDSQ 292
Db 224 STLSP-----KSLSSPDSSNSN-----SSDMTQ 246

QY 293 RPAV 296
Db 247 HPAV 250

RESULT 12
US-10-295-403-158
; Sequence 158, Application US/10295403
; Publication No. US20030101481A1
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Adam, Luc
; APPLICANT: Broun, Pierre
; APPLICANT: Pineda, Omaira
; APPLICANT: Reuber, Lynne
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Keddie, James
; APPLICANT: Zhang, James
; APPLICANT: Benito, Maria-Ines
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Fromm, Mike
; TITLE OF INVENTION: PLANT GENE SEQUENCES I
; FILE REFERENCE: MBI-0003
; CURRENT APPLICATION NUMBER: US/10/295,403
; CURRENT FILING DATE: 2002-11-15
; PRIOR APPLICATION NUMBER: US/09/394,519
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 60/101,349

```

```

; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: 60/103,312
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: 60/108,734
; PRIOR FILING DATE: 1998-11-17
; PRIOR APPLICATION NUMBER: 60/113,409
; PRIOR FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 158
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1034
US-10-295-403-158

Query Match      18.7%; Score 429.5; DB 4; Length 409;
Best Local Similarity 38.2%; Pred. No. 3.2e-20;
Matches 116; Conservative 39; Mismatches 70; Indels 79; Gaps 10;

QY 11 KFEASPA-----ESFLSAPGDNFTSLFADSTPTLNPRDMTPDVSADIDSLSVIPESQD 66
Db 8 KYENSPARTMVESFVSTPSSPHNPPLFD---NNLNPVDGFSF-----QS 48

QY 67 ABDESHSTATAPSTSEKKPVKKKSGQVLPEPKTNLPKRKRAKTEDEKEORRVERVL 126
Db 49 FDRDYNFNGSLGSLNLPKPKPIKKKSGQQLPEPKTNLPKRKRAKTODEKEORRVERVL 108

QY 127 RNRRAAQSSRRKRKEVEAL-----EKRKNELETLLINVOKTNLLILVEELNR---FRR 176
Db 109 RNRRAAQSSRRKRKEVEALEVEKRAIERKNWDLMEALADMEAKYLLQQLKCRASGYNK 168

QY 177 SSGVVTRSSSPDLSDLSQDS-----ITLSQQLFGSRDQGTWNSNPQSLMDQIMRSANPTVNP 232
Db 169 TNFLSYSDSTPDISDQSLPFTSKQLFNAQD--ELCRP---ISPOSIGPLTSRTVDP 223

QY 233 ASLSPSLPISDKFQTKDEEQADEMEQTHETKEAAAEEKSKQSRVSTDSQ 292
Db 224 STLSP-----KSLSSPDSSNSN-----SSDMTQ 246

QY 293 RPAV 296
Db 247 HPAV 250

RESULT 13
US-10-412-699B-684
; Sequence 684, Application US/10412699B
; Publication No. US20040045049A1
; GENERAL INFORMATION:
; APPLICANT: Mendel Biotechnology, Inc.
; APPLICANT: Zhang, James
; APPLICANT: Fromm, Michael E.
; APPLICANT: Heard, Jacqueline E.
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Adam, Luc J.
; APPLICANT: Broun, Pierre E.
; APPLICANT: Pineda, Omaira
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Keddie, James S.
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Samaha, Raymond R.
; APPLICANT: Pilgrim, Marsha L.
; APPLICANT: Creelman, Robert A.
; APPLICANT: DuBell, Arnold N.
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Kumamoto, Rodrick
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: Polynucleotides and Polypeptides in Plants
; FILE REFERENCE: MBI-0048CIP
; CURRENT APPLICATION NUMBER: US/10/412,699B

```

```

; CURRENT FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: 09/394,519
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 09/489,376
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: 09/506,720
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 09/533,030
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,392
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,029
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/532,591
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,648
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/713,994
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: 09/819,142
; PRIOR FILING DATE: 2001-03-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2011
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 684
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1034
US-10-412-699B-684

```

```

Query Match      18.7%; Score 429.5; DB 4; Length 409;
Best Local Similarity 38.2%; Pred. No. 3.2e-20;
Matches 116; Conservative 39; Mismatches 70; Indels 79; Gaps 10;

Qy 11 KFEASPA---ESFLSAPGDNFTSLFADSTSTLNPRDMTTPDSVADIDSRSLVSPESQD 66
Db 8 KYENSPATWTFVSTSSFNPLFD---NNLNPVDGFGSP-----QS 48

Qy 67 AEDDESHTSATPSTSEKKPVKKKSGQVLPPEKTNLPKRKRAKTEDEKEQRRVERVL 126
Db 49 FDRDYNFNGSLUGLNLPEKKIKKSGQVLPPEKTNLPKRKRAKTEDEKEQRRVERVL 108

Qy 127 RNRAAQSRERKRLEVEAL-----EKRNKELETLINQKTNLILVEELNR---PFR 176
Db 109 RNRAAQSRERKRQVEALEVEKRAIERKNWDLMLADMEAKYLLQQLKRAAGYNK 168

Qy 177 SSGVVTRSSSPDLSLQDS-----ITLSQQLFGSRDQGTNSNPEQSLMDQIMRSAANTVNP 232
Db 169 TNFLSYSDSTPDISEDSQLSPLTFKQLFNAQD---ELCRP---ISPSIGLTSRTVDP 223

Qy 233 ASLSPSLPPISDKEPQTEDEBEQADEMEQTHETKEAAAKKSKQSRVSTDTSTQ 292
Db 224 STLSP-----KSLSSPDSSNSN-----SSDWTQ 246

Qy 293 RPAV 296
Db 247 HPAV 250

```

```

RESULT 14
US-09-816-277-19
; Sequence 19, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells

```

```

; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-09-816-277-19

Query Match      18.6%; Score 426.5; DB 3; Length 342;
Best Local Similarity 31.7%; Pred. No. 4.1e-20;
Matches 145; Conservative 44; Mismatches 125; Indels 143; Gaps 20;

Qy 15 SPAESFLSAPGDNFTSLFADSTPS---TLAPRDMTTPDSVADIDSRSLVSPESQDAEDD 70
Db 7 SPVDSLAGSP-----TPELPLUTVSPADTSLDSS-----35

Qy 71 ESHSTSATPSTSEKKPVKKKSGQVLPPEKTNLPKRKRAKTEDEKEQRRVERVLNR 130
Db 36 ---SVQAGETKAEKKPVKKKSGQELPVKTNLPKRKRAKTEDEKEQRRVERVLNR 92

Qy 131 AAQSRERKRLEVEALEKRNKELETLINQKTNLILVEELNRPRSSGVVTRSSPLDS 190
Db 93 AAQTSRERKRLEMEKLENEK-----IQMEQQNQFLQLRSLQMEANRLN-----137

Qy 191 LQDSITLSQQLFGSRDQGTNSNPEQSLMDQIMRSAANTVNPASLSPSLPISDKKEFQTK 250
Db 138 -QQVAQLSAEVRGSR-GNT---PKPG-----SPVSASPLTLPFLFKQ-----174

Qy 251 EEDS---EQADEDEMEQTHETKEAAAKKSKQSRVSTDTSTQRPVSIIGG---DAA 303
Db 175 ERDIPLERIFPPPTSIDYSPTLRPSTLAE-----SSDVTQHPAVSVAGLEGE 226

Qy 304 VPVFSDDAGANCLGDPVHQDGGPFSGHSGLSAALDADRY--LLESQLLASPNASTVD 361
Db 227 LSLF--DVGSN---PEPHAADD-----LAAPLSDDDFHRLFNVDSPVGSDDSVLE 271

Qy 362 D---DVLGDSAACTNPLPSDYDFDINDFLTDHANAAVDIVAASNYAAADRELDLEI 417
Db 272 DGFAPDVLDDGDLSPF---PFDSMVDFD-----PESVGFEG 304

Qy 418 HDPENQIPSRHSIQ-----QPQSGASSHGCDGDIAGV 450
Db 305 IEPHGLPDETSRGTSSVQPSLGASTRCDGQGIAG 341

```

```

RESULT 15
US-10-663-450-19
; Sequence 19, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Aspergillus niger

```

Sat Nov 26 13:50:38 2005

US-10-663-450-19

```
Query Match      18.6%; Score 426.5; DB 4; Length 342;
Best Local Similarity 31.7%; Pred. No. 4.1e-20;
Matches 145; Conservative 44; Mismatches 125; Indels 143; Gaps 20;

Qy      15 SPAESFLSAPGDNFTSLFADSTPS---TLNPRDMTPTDSVADIDSRLSVIPESQDAEDD 70
Db      7 SPVDSLAGSP-----TPELPLTVSPADTSLDDS-----35

Qy      71 ESHSTASATPSTSKKPVKKKSGQVLPKPTNLPPKRAKTEDEKEORVERVLNRR 130
Db      36 ---SVQAGETKAEEKKPVKKKSGQELPVKTNLPPKRAKTEDEKEQRRIERVLRNA 92

Qy      131 AAQSSRRERKLEVALEKRNKELETLLINVOKTLLIIVVEELNRRFRSSGVVTRSSPLDS 190
Db      93 AAQTSRRERKLEMEKLENEK-----IQMEQQNQFLLQRLSQMEANNRLN-----137

Qy      191 LQDSITLSQQLFGSRDGTMSNPBQSLMDQIMRSAANPTVNPASLSPSLPISDKFQTK 250
Db      138 -QQVAQLSAEVRGSR-GWT---PKPG-----SPVSASPTLTPTLFKQ----174

Qy      251 EEDE---EQADEDEMEQTWHTKEAATAAANKSKQSRVSTDTSTORPAVSIGG-----DAA 303
Db      175 ERDEIPLERIIPPTPSITDYSPTLRPSTLAE-----SSDVTQHPAVSVAGLEGESA 226

Qy      304 VPVFSDDAGANCLGLDPVHQDDGPFSIGHSGLSAALDADRY--LLESQLLASPNASTVD 361
Db      227 LSLF--DVGSN---PEPHAADD-----LAPLSDDDFHRLFNVDSPVGSDDSVLE 271

Qy      362 D----DYLAGDSACFTNPLFSDYDFDINDFLTDANHAAYDIVAASNAAADRELDLEI 417
Db      272 DGFAFDVLDGGLSAF--PFDSMWDFD-----PESVGFEG 304

Qy      418 HDPENQIPSRHSIQ-----QPSGASHGCGDDGGIANG 450
Db      305 IEPHGLPDETSRQTSSVQPSLGASTSRCGQGIAG 341
```

Search completed: November 23, 2005, 03:32:19
Job time : 148.003 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:18:52 ; Search time 0.861751 Seconds
(without alignments)
239.171 Million cell updates/sec

Title: US-10-663-450-60

Perfect score: 342

Sequence: 1 KSTLPKRAKTKKEQR.....CSLLENLNSVNLKLADE 68

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 17545 seqs, 3030971 residues

Total number of hits satisfying chosen parameters: 17545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/prodata/2/pubaa/US10_NEW_PUB_PEP.*
- 2: /cgn2_6/prodata/2/pubaa/US06_NEW_PUB_PEP.*
- 3: /cgn2_6/prodata/2/pubaa/US07_NEW_PUB_PEP.*
- 4: /cgn2_6/prodata/2/pubaa/US08_NEW_PUB_PEP.*
- 5: /cgn2_6/prodata/2/pubaa/US09_NEW_PUB_PEP.*
- 6: /cgn2_6/prodata/2/pubaa/PCT_NEW_PUB_PEP.*
- 7: /cgn2_6/prodata/2/pubaa/US11_NEW_PUB_PEP.*
- 8: /cgn2_6/prodata/2/pubaa/US60_NEW_PUB_PEP.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87.5	25.6	135	1	US-10-821-234-1037
2	72	21.1	454	7	US-11-169-013-1
3	69	20.2	376	1	US-10-821-234-1037
4	63	18.4	579	1	US-10-821-234-874
5	62.5	18.3	431	7	US-11-169-013-2
6	61.5	18.0	86	1	US-10-667-295-128
7	61.5	18.0	97	1	US-10-667-295-127
8	61.5	18.0	431	1	US-10-821-234-1285
9	60	17.3	383	1	US-10-793-626-332
10	59	17.3	453	1	US-10-131-826A-270
11	59	17.3	457	1	US-10-821-234-1185
12	58.5	17.1	1388	1	US-10-821-234-1143
13	56	16.4	625	1	US-10-310-386-6
14	56	16.4	776	1	US-10-821-234-1171
15	55.5	16.2	82	1	US-10-667-295-69
16	55.5	16.2	90	1	US-10-667-295-67
17	55.5	16.2	185	1	US-10-821-234-1130
18	55.5	16.2	586	1	US-10-131-826A-46
19	55	16.1	112	7	US/11/020
20	55	16.1	284	1	US-10-821-234-1632
21	54.5	15.9	639	1	US-10-821-234-907
22	54.5	15.9	2432	1	US-10-821-234-899
23	54	15.8	401	1	US-10-821-234-881
24	53.5	15.6	88	1	US-10-667-295-190
25	53.5	15.6	1614	1	US-10-821-234-903

26	53	15.5	171	1	US-10-821-234-994	Sequence 994, App
27	52.5	15.4	128	7	US-11-105-864-4	Sequence 4, Appli
28	52.5	15.4	578	1	US-10-821-234-1039	Sequence 1039, Ap
29	52	15.2	267	1	US-10-793-626-882	Sequence 882, App
30	52	15.2	400	7	US-11-018-018-7	Sequence 7, Appli
31	52	15.2	400	7	US-11-047-757-7	Sequence 7, Appli
32	52	15.2	747	7	US-11-182-592-2	Sequence 2, Appli
33	51.5	15.1	163	1	US-10-689-742-42	Sequence 42, Appl
34	51.5	15.1	475	1	US-10-793-626-2076	Sequence 2076, Ap
35	51.5	15.1	525	1	US-10-793-626-1292	Sequence 1292, Ap
36	51.5	15.1	559	1	US-10-821-234-1513	Sequence 1513, Ap
37	51.5	15.1	719	1	US-10-793-626-1548	Sequence 1548, Ap
38	51	14.9	322	1	US-10-793-626-1990	Sequence 1990, Ap
39	51	14.9	877	7	US-11-074-176-322	Sequence 322, App
40	51	14.9	883	7	US-11-074-176-88	Sequence 88, Appl
41	51	14.9	1616	1	US-10-821-234-1497	Sequence 1497, Ap
42	50.5	14.8	122	1	US-10-667-295-177	Sequence 177, App
43	50.5	14.8	1059	7	US/11/062	Sequence 5, Appli
44	50.5	14.8	1084	7	US/11/062	Sequence 8, Appli
45	50	14.6	410	1	US-10-821-234-1180	Sequence 1180, Ap

ALIGNMENTS

RESULT 1

US-10-821-234-1037
; Sequence 1037, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; PRIOR FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1037
; LENGTH: 135
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1037

Query Match 25.6%; Score 87.5; DB 1; Length 135;
Best Local Similarity 34.5%; Pred. No. 0.002;
Matches 20; Conservative 16; Mismatches 21; Indels 1; Gaps 1;
QY 7 RKRATKEKEQRRIERILNRRAHQSRKKRLHLQYLERKGLLENLNSVNLK 64
DB 67 RSFQLAEEATRKRELLMKNRREARCKRKKEYKCLNRAVAVLEN-QNKTLEBL 123

RESULT 2

US-11-169-013-1
; Sequence 1, Application US/11169013
; Publication No. US20050244971A1
; GENERAL INFORMATION:
; APPLICANT: Korea Kumho Petrochemical Co., Ltd.
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH ENHANCED STRESS TOLERANCE
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/11/169,013
; CURRENT FILING DATE: 2005-06-29
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 1
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana

```
US-11-169-013-1
Query Match      21.1%; Score 72; DB 7; Length 454;
Best Local Similarity 32.7%; Pred. No. 0.33;
Matches 16; Conservative 14; Mismatches 19; Indels 0; Gaps 0;

QY 7 RRAKTKKEEQRRIERILNRRAAHQSRKRLHLQYLERKCSLLENL 55
Db 362 RKTGAVLEKVIERRQRMKIKNRESAARSRAKQAYTMELEAEIAQLKEL 410

RESULT 3
US-10-971-994-2
; Sequence 2, Application US/10971994
; Publication No. US20050250182A1
; GENERAL INFORMATION:
; APPLICANT: University of Michigan et al.
; APPLICANT: Kaufman, Randal
; APPLICANT: Kyungo, Lee
; APPLICANT: Mori, Kazutos
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, AND METHODS FOR
; TITLE OF INVENTION: MODULATING THE UNFOLDED PROTEIN RESPONSE
; FILE REFERENCE: UMW-2246
; CURRENT APPLICATION NUMBER: US/10/971,994
; CURRENT FILING DATE: 2004-10-21
; PRIOR APPLICATION NUMBER: PCT/2003/012640
; PRIOR FILING DATE: 2003-04-22
; PRIOR APPLICATION NUMBER: 60/375,098
; PRIOR FILING DATE: 2002-04-22
; PRIOR APPLICATION NUMBER: 60/374,980
; PRIOR FILING DATE: 2002-04-23
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-971-994-2

Query Match      20.2%; Score 69; DB 1; Length 376;
Best Local Similarity 33.3%; Pred. No. 0.55;
Matches 20; Conservative 12; Mismatches 18; Indels 10; Gaps 2;

QY 7 RRAKTKKEEQRRIERILNRRAAHQSRKRLHLQYLERKCSLLENL 56
Db 59 RKRQRLTHLSPKALRRKIKRVAQAQTARDKAKMSLEQQVVDLEENQKLLLENQL 118

RESULT 4
US-10-821-234-874
; Sequence 874, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 874
; LENGTH: 579
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-874

Query Match      18.4%; Score 63; DB 1; Length 579;
Best Local Similarity 45.5%; Pred. No. 3.8;

us-10-663-450-60.rapbn

Matches 15; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

QY 7 RRAKTKKEEQRRIERILNRRAAHQSRKRLHLQYLERKCSLLENL 39
Db 96 RKRKSKERKRRDRERKSKSRKRSKR 128

RESULT 5
US-11-169-013-2
; Sequence 2, Application US/11169013
; Publication No. US20050244971A1
; GENERAL INFORMATION:
; APPLICANT: Korea Kumho Petrochemical Co., Ltd.
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH ENHANCED STRESS TOLERANCE
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/11/169,013
; CURRENT FILING DATE: 2005-06-29
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 2
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-11-169-013-2

Query Match      18.3%; Score 62.5; DB 7; Length 431;
Best Local Similarity 34.0%; Pred. No. 3.1;
Matches 16; Conservative 13; Mismatches 17; Indels 1; Gaps 1;

QY 8 RRAKTKKEEQRRIERILNRRAAHQSRKRLHLQYLERKCSLLENL 53
Db 341 RRSNTGLEKVIERRQRMKIKNRESAARSRAKQAYTMELEAEIAQLK 387

RESULT 6
US-10-667-295-128
; Sequence 128, Application US/10667295
; Publication No. US20050257293A1
; GENERAL INFORMATION:
; APPLICANT: Mascia, Peter
; TITLE OF INVENTION: BIOLOGICAL CONTAINMENT SYSTEM
; FILE REFERENCE: 11696-047001
; CURRENT APPLICATION NUMBER: US/10/667,295
; CURRENT FILING DATE: 2003-09-17
; PRIOR APPLICATION NUMBER: US 60/411,823
; PRIOR FILING DATE: 2002-09-17
; NUMBER OF SEQ ID NOS: 263
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 128
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(86)
; OTHER INFORMATION: Ceres Seq. ID no. 12438738
US-10-667-295-128

Query Match      18.0%; Score 61.5; DB 1; Length 86;
Best Local Similarity 32.3%; Pred. No. 0.6;
Matches 21; Conservative 14; Mismatches 25; Indels 5; Gaps 3;

QY 8 RRAKTKKEEQRRIERILNRRAAHQSRKRLHLQYLERKCSLLENL 62
Db 14 RAARSKSEERCKIESIIISQEFRNAEDVDKQNFYIEYLLRGNKQLQKQKPGTTGLS 73

QY 63 KLADH 67
Db 74 SLQLH 78

RESULT 7
US-10-667-295-127
```



```

; Sequence 127, Application US/10567295
; Publication No. US20050257293A1
; GENERAL INFORMATION:
; APPLICANT: Mascia, Peter
; TITLE OF INVENTION: BIOLOGICAL CONTAINMENT SYSTEM
; FILE REFERENCE: 11696-047001
; CURRENT APPLICATION NUMBER: US/10/667,295
; CURRENT FILING DATE: 2003-09-17
; PRIOR APPLICATION NUMBER: US 60/411,823
; PRIOR FILING DATE: 2002-09-17
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 127
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(97)
; OTHER INFORMATION: Ceres Seq. ID no. 12438737
US-10-667-295-127

Query Match      18.0%; Score 61.5; DB 1; Length 97;
Best Local Similarity 32.3%; Pred. No. 0.69;
Matches 21; Conservative 14; Mismatches 25; Indels 5; Gaps 3;

QY 8 KRAKTKKEEKQRIIRLN--RRAHQSRKKRLHLQYLERKCS-LLENLLN--SYNLE 62
Db 25 RAARSKSEERCKIESIISQEFRNAEDVRKNFLYIEYLLRRGNQLDQKPKPGTGLS 84
QY 63 KLAHD 67
Db 85 SLQLH 89

RESULT 8
US-10-821-234-1285
; Sequence 1285, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1285
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1285

Query Match      18.0%; Score 61.5; DB 1; Length 431;
Best Local Similarity 32.2%; Pred. No. 3.9;
Matches 19; Conservative 14; Mismatches 21; Indels 5; Gaps 2;

QY 8 KRAKTK---EKEQRIIRLNRAAHQSRKKRLHLQYLERKCSLLENLLNSYNLE 62
Db 176 RRGKVLQSPSEBKRIIR-ERNKAAKCRNRRLDTTLQAEYTDQLEKSAQTE 233

RESULT 9
US-10-793-626-3332
; Sequence 3332, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN

```

; Remaining Prior Application data removed - See File Wrapper or PALM.

```
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 270
; LENGTH: 453
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-826A-270

Query Match      17.3%; Score 59; DB 1; Length 453;
Best Local Similarity 26.7%; Pred. No. 7.6;
Matches 20; Conservative 15; Mismatches 20; Indels 20; Gaps 4;

QY 4 LPPRKRA-----KTKEEKQRIERILRNRAAHQSR--EKRLHLQYLERKCSLLE 53
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 138 LAPESRVDPVLSLQTLKAQNSRIQLF--HKVAQQQRHLEKQHLRIQHLSQFGLLD 195
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 54 NLLNSVNLEKLADHE 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 196 H-----KHLDE 202
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 11
US-10-821-234-1185
; Sequence 1185, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Presclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt SEQ_genes Version 1.0
; SEQ ID NO 1185
; LENGTH: 457
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1185
```

```
Query Match      17.3%; Score 59; DB 1; Length 457;
Best Local Similarity 26.7%; Pred. No. 7.6;
Matches 20; Conservative 15; Mismatches 20; Indels 20; Gaps 4;

QY 4 LPPRKRA-----KTKEEKQRIERILRNRAAHQSR--EKRLHLQYLERKCSLLE 53
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 142 LAPESRVDPVLSLQTLKAQNSRIQLF--HKVAQQQRHLEKQHLRIQHLSQFGLLD 199
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 54 NLLNSVNLEKLADHE 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 200 H-----KHLDE 206
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 12
US-10-821-234-1143
; Sequence 1143, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Presclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
```

```
; SOFTWARE: pt SEQ_genes Version 1.0
; SEQ ID NO 1143
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1143

Query Match      17.1%; Score 58.5; DB 1; Length 1388;
Best Local Similarity 29.2%; Pred. No. 31;
Matches 19; Conservative 10; Mismatches 33; Indels 3; Gaps 1;

QY 4 LPPRKRAKTKEEKQRIERILRNRAAHQSRKKRLHLQYLERKCSLLENLLNSVNLK 63
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 543 IKPAHILQKEEQHLAVTAVLKNSRKEHQRIARR---QTTEERKERLESNIQREKEE 599
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 64 LADHE 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 600 LEQRE 604
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 13
US-10-510-386-6
; Sequence 6, Application US/10510386
; Publication No. US20050244922A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Clausen, Ib Groth
; APPLICANT: Jorgensen, Steen Troels
; APPLICANT: Olsen, Peter Bjarke
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: Improved Bacillus Host Cell
; FILE REFERENCE: 10294,204-US
; CURRENT APPLICATION NUMBER: US/10/510,386
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 6
; LENGTH: 625
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-10-510-386-6
```

```
Query Match      16.4%; Score 56; DB 1; Length 625;
Best Local Similarity 61.9%; Pred. No. 23;
Matches 13; Conservative 4; Mismatches 2; Indels 2; Gaps 1;

QY 8 KRAKTKEEKQRIERILRN 28
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 295 KKAKTKEEK--KKIEKLSNR 313
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 14
US-10-821-234-1171
; Sequence 1171, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Presclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt SEQ_genes Version 1.0
; SEQ ID NO 1171
; LENGTH: 776
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1171
```

Query Match 16.4%; Score 56; DB 1; Length 776;
Best Local Similarity 27.3%; Pred. No. 29;
Matches 18; Conservative 11; Mismatches 21; Indels 16; Gaps 2;
Qy 14 BEKEQRRIERI-----LRNRRAAHQGREKKRLHLQYLERKCSLLENLLNSVNL 62
Db 479 QAEQDKIEVRDAMENEMRTQLRRQAAATDHLRDVLRVQEQELKSEFEQNL-----SE 533
Qy 63 KLDHDE 68
Db 534 KLSEQE 539

RESULT 15
US-10-667-295-69
; Sequence 69, Application US/10667295
; Publication No. US20050257293A1
; GENERAL INFORMATION:
; APPLICANT: Macchia, Peter
; TITLE OF INVENTION: BIOLOGICAL CONTAINMENT SYSTEM
; FILE REFERENCE: 11696-047001
; CURRENT APPLICATION NUMBER: US/10/667,295
; CURRENT FILING DATE: 2003-09-17
; PRIOR APPLICATION NUMBER: US 60/411,823
; PRIOR FILING DATE: 2002-09-17
; NUMBER OF SEQ ID NOS: 263
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(82)
; OTHER INFORMATION: Ceres Seq. ID no. 12561624
US-10-667-295-69

Query Match 16.2%; Score 55.5; DB 1; Length 82;
Best Local Similarity 30.8%; Pred. No. 2.4;
Matches 12; Conservative 13; Mismatches 13; Indels 1; Gaps 1;
Qy 8 KRAKTKKEQRIERIILNRRAAHQGREKKRLHLQYLE 46
Db 14 RAARSRPIDRRRIEMIV-STEFHNSKEVDKKNFYIE 51

Search completed: November 23, 2005, 03:32:38
Job time : 0.861751 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:12:56 ; Search time 22.0138 Seconds
(without alignments)
1290.661 Million cell updates/sec

Title: US-10-663-450-60

Perfect score: 342

Sequence: 1 KSTLPPRKAKTKEKEQRR.....CSLLENLNSVNLEKLADHE 68

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	342	100.0	68	3	US-09-816-277-60
2	342	100.0	68	4	US-10-663-450-60
3	342	100.0	200	4	US-10-369-493-21867
4	196	57.3	451	3	US-09-816-277-2
5	196	57.3	451	3	US-09-816-277-5
6	196	57.3	451	4	US-10-663-450-2
7	196	57.3	451	4	US-10-663-450-5
8	178.5	52.2	409	3	US-09-533-029-104
9	178.5	52.2	409	4	US-10-295-403-158
10	178.5	51.6	342	3	US-09-816-277-19
11	176.5	51.6	342	3	US-09-816-277-16
12	176.5	51.6	342	4	US-10-663-450-19
13	176.5	51.6	349	3	US-09-816-277-4
14	176.5	51.6	349	3	US-09-816-277-6
15	176.5	51.6	349	4	US-10-663-450-4
16	176.5	51.6	349	4	US-10-663-450-6
17	176.5	51.6	386	3	US-09-816-277-16
18	176.5	51.6	386	4	US-10-663-450-16
19	171.5	50.1	174	4	US-10-767-701-61238
20	124.5	36.4	168	3	US-09-934-455-102
21	124.5	36.4	168	4	US-10-225-066A-206
22	124.5	36.4	168	4	US-10-374-780A-2356
23	124.5	36.4	168	5	US-10-732-923-13692
24	124.5	36.4	168	5	US-10-225-066A-206
25	122	35.7	203	4	US-10-437-963-151695
26	122	35.7	203	5	US-10-732-923-13694
27	122	35.7	203	5	US-10-732-923-13695

US-09-816-277-60
; Sequence 60, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816, 277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534, 692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 60
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-816-277-60
Query Match 100.0%; Score 342; DB 3; Length 68;
Best Local Similarity 100.0%; Pred. No. 2.2e-27; Mismatches 0; Indels 0; Gaps 0;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KSTLPPRKAKTKEKEQRRRIILNRRRAAHOSREKRLHLQYLERKCSLLENLNSVN 60
DB 1 KSTLPPRKAKTKEKEQRRRIILNRRRAAHOSREKRLHLQYLERKCSLLENLNSVN 60
QY 61 LEKLADHE 68
DB 61 LEKLADHE 68
RESULT 2
US-10-663-450-60
; Sequence 60, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells

ALIGNMENTS

RESULT 1
US-09-816-277-60
; Sequence 60, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816, 277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534, 692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 60
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-816-277-60
Query Match 100.0%; Score 342; DB 3; Length 68;
Best Local Similarity 100.0%; Pred. No. 2.2e-27; Mismatches 0; Indels 0; Gaps 0;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KSTLPPRKAKTKEKEQRRRIILNRRRAAHOSREKRLHLQYLERKCSLLENLNSVN 60
DB 1 KSTLPPRKAKTKEKEQRRRIILNRRRAAHOSREKRLHLQYLERKCSLLENLNSVN 60
QY 61 LEKLADHE 68
DB 61 LEKLADHE 68
RESULT 2
US-10-663-450-60
; Sequence 60, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells

```
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 60
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-663-450-60

Query Match      100.0%; Score 342; DB 4; Length 68;
Best Local Similarity 100.0%; Pred. No. 2.2e-27;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSTLPPRKAKTKEEKEQRRIERILNRRRAAHOQSREKKRLHLQYLERKCSLLENLNSVN 60
   |||||
Db 1 KSTLPPRKAKTKEEKEQRRIERILNRRRAAHOQSREKKRLHLQYLERKCSLLENLNSVN 60
   |||||
QY 61 LEKLADHE 68
   |||||
Db 61 LEKLADHE 68
   |||||

RESULT 3
US-10-369-493-21867
; Sequence 21867, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 21867
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-21867

Query Match      100.0%; Score 342; DB 4; Length 200;
Best Local Similarity 100.0%; Pred. No. 6.7e-27;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSTLPPRKAKTKEEKEQRRIERILNRRRAAHOQSREKKRLHLQYLERKCSLLENLNSVN 60
   |||||
Db 23 KSTLPPRKAKTKEEKEQRRIERILNRRRAAHOQSREKKRLHLQYLERKCSLLENLNSVN 82
   |||||
QY 61 LEKLADHE 68
   |||||
Db 83 LEKLADHE 90
   |||||

RESULT 4
US-09-816-277-2
; Sequence 2, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
```

```
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-09-816-277-2

Query Match      57.3%; Score 196; DB 3; Length 451;
Best Local Similarity 61.9%; Pred. No. 1.1e-11;
Matches 39; Conservative 13; Mismatches 9; Indels 2; Gaps 1;

QY 1 KSTLPPRKAKTKEEKEQRRIERILNRRRAAHOQSREKKRLHLQYLERKCSLLENLNSVN 60
   |||||
Db 102 KTNLPPRKAKTDEKEQRRVERVLNRRRAAQSRRERKELEVEALEKRNKELETL--IN 159
   |||||
QY 61 LEK 63
   ||
Db 160 VQK 162
   ||

RESULT 5
US-09-816-277-5
; Sequence 5, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-09-816-277-5

Query Match      57.3%; Score 196; DB 3; Length 451;
Best Local Similarity 61.9%; Pred. No. 1.1e-11;
Matches 39; Conservative 13; Mismatches 9; Indels 2; Gaps 1;

QY 1 KSTLPPRKAKTKEEKEQRRIERILNRRRAAHOQSREKKRLHLQYLERKCSLLENLNSVN 60
   |||||
Db 102 KTNLPPRKAKTDEKEQRRVERVLNRRRAAQSRRERKELEVEALEKRNKELETL--IN 159
   |||||
QY 61 LEK 63
   ||
Db 160 VQK 162
   ||

RESULT 6
US-10-663-450-2
; Sequence 2, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
```

```
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-663-450-2

Query Match          57.3%; Score 196; DB 4; Length 451;
Best Local Similarity 61.9%; Pred. No. 1.1e-11;
Matches 39; Conservative 13; Mismatches 9; Indels 2; Gaps 1;

Qy 1 KSTLPPRKRAKTEKEQRRIILNRRAAHSREKKRLHLQYLRKCSLLENLNSVN 60
Db 102 KTNLPPRKRAKTEDEKEQRVVERVLRNRAAQSRRERKRLEVEALEKRNKELETL--IN 159

Qy 61 LEK 63
Db 160 VQK 162

RESULT 7
US-10-663-450-5
; Sequence 5, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-663-450-5

Query Match          57.3%; Score 196; DB 4; Length 451;
Best Local Similarity 61.9%; Pred. No. 1.1e-11;
Matches 39; Conservative 13; Mismatches 9; Indels 2; Gaps 1;

Qy 1 KSTLPPRKRAKTEKEQRRIILNRRAAHSREKKRLHLQYLRKCSLLENLNSVN 60
Db 102 KTNLPPRKRAKTEDEKEQRVVERVLRNRAAQSRRERKRLEVEALEKRNKELETL--IN 159

Qy 61 LEK 63
Db 160 VQK 162

RESULT 8
US-09-533-029-104
; Sequence 104, Application US/09533029
; Publication No. US20030046723A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Broun, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Keddle, James
; APPLICANT: Pineda, Omaira
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pilgrim, Mareha
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Reuber, Lynne
; TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES
; FILE REFERENCE: MBI-010
; CURRENT APPLICATION NUMBER: US/09/533,029
; CURRENT FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 60/125,814
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 104
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1034
US-09-533-029-104

Query Match          52.2%; Score 178.5; DB 3; Length 409;
Best Local Similarity 56.5%; Pred. No. 5.0e-10;
Matches 39; Conservative 14; Mismatches 13; Indels 3; Gaps 2;

Qy 1 KSTLPPRKRAKTEKEQRRIILNRRAAHSREKKRLHLQYLRKCSLLENLNSVN 60
Db 84 KTNLPPRKRAKTEDEKEQRVVERVLRNRAAQSRRERKRQVEALEVEKRAIER--KNMD 141

Qy 61 LE-KLADHE 68
Db 142 LEMRLADME 150

RESULT 9
US-10-295-403-158
; Sequence 158, Application US/10295403
; Publication No. US20030101481A1
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Adam, Luc
; APPLICANT: Broun, Pierre
; APPLICANT: Pineda, Omaira
; APPLICANT: Reuber, Lynne
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Keddle, James
; APPLICANT: Zhang, James
; APPLICANT: Benito, Maria-Ines
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Fromm, Mike
; TITLE OF INVENTION: PLANT GENE SEQUENCES I
; FILE REFERENCE: MBI-0003
; CURRENT APPLICATION NUMBER: US/10/295,403
; CURRENT FILING DATE: 2002-11-15
; PRIOR APPLICATION NUMBER: US/09/394,519
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 60/101,349
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: 60/103,312
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: 60/108,734
; PRIOR FILING DATE: 1998-11-17
; PRIOR APPLICATION NUMBER: 60/113,409
```

```
; PRIOR FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 158
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1034
US-10-295-403-158

Query Match          52.2%; Score 178.5; DB 4; Length 409;
Best Local Similarity 56.5%; Pred. No. 5.8e-10;
Matches 39; Conservative 14; Mismatches 13; Indels 3; Gaps 2;

QY 1 KSTLPKRAKTKKEQRRRIILNRRRAAHQSREKKRLHLQYLERKCSLLENLNSYN 60
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 84 KTNLPKRAKTKDEKQRRVERVLNRRRAAQSRRERKQVEALEVEKRAIER--KNMD 141
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 LE-KLADHE 68
   ||:|||||
Db 142 LEMRLADME 150

RESULT 10
US-10-412-699B-684
; Sequence 684, Application US/10413699B
; Publication No. US20040045049A1
; GENERAL INFORMATION:
; APPLICANT: Mendel Biotechnology, Inc.
; APPLICANT: Zhang, James
; APPLICANT: Fromm, Michael E.
; APPLICANT: Heard, Jacqueline E.
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Adam, Luc J.
; APPLICANT: Broun, Pierre E.
; APPLICANT: Pineda, Omaira
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Keddle, James S.
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Samaha, Raymond R.
; APPLICANT: Pilgrim, Marsha L.
; APPLICANT: Creelman, Robert A.
; APPLICANT: DuBell, Arnold N.
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Kumimoto, Roderick
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: Polynucleotides and Polypeptides in Plants
; FILE REFERENCE: MBI-0048CIP
; CURRENT APPLICATION NUMBER: US/10/412,699B
; CURRENT FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: 09/394,519
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 09/489,376
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: 09/506,720
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 09/533,030
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,392
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,029
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/532,591
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,648
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/713,994
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: 09/819,142
; PRIOR FILING DATE: 2001-03-27
; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 2011
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 684
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1034
US-10-412-699B-684

Query Match          52.2%; Score 178.5; DB 4; Length 409;
Best Local Similarity 56.5%; Pred. No. 5.8e-10;
Matches 39; Conservative 14; Mismatches 13; Indels 3; Gaps 2;

QY 1 KSTLPKRAKTKKEQRRRIILNRRRAAHQSREKKRLHLQYLERKCSLLENLNSYN 60
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 84 KTNLPKRAKTKDEKQRRVERVLNRRRAAQSRRERKQVEALEVEKRAIER--KNMD 141
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 LE-KLADHE 68
   ||:|||||
Db 142 LEMRLADME 150

RESULT 11
US-09-816-277-19
; Sequence 19, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GCS90-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-09-816-277-19

Query Match          51.6%; Score 176.5; DB 3; Length 342;
Best Local Similarity 55.9%; Pred. No. 7.7e-10;
Matches 38; Conservative 12; Mismatches 17; Indels 1; Gaps 1;

QY 1 KSTLPKRAKTKKEQRRRIILNRRRAAHQSREKKRLHLQYLERKCSLLENLNSYN 60
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 64 KTNLPKRAKTKDEKQRRRIILNRRRAAQSRRERKQVEALEVEKRAIER--KNMD 122
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 LE-KLADHE 68
   ||:|||||
Db 123 LORLSQME 130

RESULT 12
US-10-663-450-19
; Sequence 19, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
```



```
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-10-663-450-19

Query Match
Best Local Similarity 51.6%; Score 176.5; DB 4; Length 342;
Matches 38; Conservative 12; Mismatches 17; Indels 1; Gaps 1;

QY 1 KSTLPPRKAKTKKEQRRRIILNRRAAHQSRKKRLHLYLQYLRKCSLLENLNSVN 60
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 64 KTNLPKRKAKTEDEKEQRRRIILNRRAAAQTSRKRKLEMEKLESEKIDMEQ-QNQFL 122
   |:|:|:|

QY 61 LEKLADHE 68
   |:|:|:|
Db 123 LORLSQME 130

RESULT 13
US-09-816-277-4
; Sequence 4, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-816-277-4

Query Match
Best Local Similarity 51.6%; Score 176.5; DB 3; Length 349;
Matches 39; Conservative 11; Mismatches 17; Indels 1; Gaps 1;

QY 1 KSTLPPRKAKTKKEQRRRIILNRRAAHQSRKKRLHLYLQYLRKCSLLENLNSVN 60
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 70 KTNLPKRKAKTEDEKEQRRRIILNRRAAAQTSRKRKLEMEKLESEKIDMEQ-QNQFL 128
   |:|:|:|

QY 61 LEKLADHE 68
   |:|:|:|
Db 129 LORLAQME 136

RESULT 14
US-09-816-277-6
; Sequence 6, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-663-450-4

Query Match
Best Local Similarity 51.6%; Score 176.5; DB 4; Length 349;
Matches 39; Conservative 11; Mismatches 17; Indels 1; Gaps 1;

QY 1 KSTLPPRKAKTKKEQRRRIILNRRAAHQSRKKRLHLYLQYLRKCSLLENLNSVN 60
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 70 KTNLPKRKAKTEDEKEQRRRIILNRRAAAQTSRKRKLEMEKLESEKIDMEQ-QNQFL 128
   |:|:|:|

QY 61 LEKLADHE 68
   |:|:|:|
Db 129 LORLAQME 136

Search completed: November 23, 2005, 03:32:22
Job time : 23.0138 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2005, 15:43:54 ; Search time 4 Seconds
(without alignments)
18.077 Million cell updates/sec

Title: US-10-663-450-6_COPY_53_116

Perfect score: 324

Sequence: 1 KPAKRRKSGWQLPVKTN.....RAAAQTSRRKRLEMEKLS 64

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 8323 seqs, 1129788 residues

Total number of hits satisfying chosen parameters: 8323

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New.*

- 1: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pap.*
- 2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pap.*
- 3: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pap.*
- 4: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pap.*
- 5: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pap.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pap.*
- 7: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pap.*
- 8: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91	28.1	376	1	US-10-971-994-2
2	69.5	21.5	605	1	US-10-131-826A-160
3	65.5	20.2	431	7	US-11-169-013-2
4	64	19.8	454	7	US-11-169-013-1
5	63.5	19.6	879	7	US-11-022-562-340
6	59	18.2	400	7	US-11-018-018-7
7	56	17.3	134	1	US-10-329-472-2
8	55.5	17.1	194	7	US-11-074-176-82
9	55	17.0	625	1	US-10-510-386-6
10	54.5	16.8	615	1	US-10-982-545-14
11	54	16.7	21	1	US-10-995-951A-33
12	54	16.7	903	1	US-10-689-742-142
13	53	16.4	188	1	US-10-131-826A-2
14	51.5	15.9	860	7	US-11-022-562-217
15	51	15.7	312	1	US-10-986-501-124
16	51	15.7	747	7	US-11-182-592-2
17	51	15.7	804	1	US-10-467-962B-57
18	51	15.7	1302	7	US-11-004-057-6
19	51	15.7	1493	7	US-11-004-057-4
20	51	15.7	1493	7	US-11-004-057-21
21	50.5	15.6	128	7	US-11-105-864-4
22	50.5	15.6	240	1	US-10-689-742-210
23	50.5	15.6	1618	1	US-10-984-645-2
24	49.5	15.3	718	1	US-10-467-962B-29
25	49.5	15.3	718	1	US-10-467-962B-65

26	49	15.1	1023	1	US-10-131-826A-200	Sequence 200, Appl
27	48.5	15.0	180	1	US-10-689-742-84	Sequence 84, Appl
28	48.5	15.0	327	7	US-11-090-878-16	Sequence 16, Appl
29	48	14.8	2897	1	US-10-499-715-2	Sequence 2, Appl
30	47.5	14.7	234	1	US-10-997-697-47	Sequence 47, Appl
31	47.5	14.7	453	1	US-10-131-826A-270	Sequence 270, Appl
32	47.5	14.7	476	1	US-10-997-697-29	Sequence 29, Appl
33	47.5	14.7	526	1	US-10-997-697-28	Sequence 28, Appl
34	47.5	14.7	790	1	US-10-131-826A-204	Sequence 204, Appl
35	47.5	14.7	1267	7	US-11-109-156-35	Sequence 35, Appl
36	47	14.5	127	1	US-10-997-437A-8	Sequence 8, Appl
37	47	14.5	221	1	US-10-636-716-60	Sequence 60, Appl
38	47	14.5	877	7	US-11-074-176-322	Sequence 322, Appl
39	47	14.5	883	7	US-11-074-176-88	Sequence 88, Appl
40	46.5	14.4	26	7	US-11-019-894A-20	Sequence 20, Appl
41	46.5	14.4	197	1	US-10-986-501-296	Sequence 296, Appl
42	46.5	14.4	435	1	US-10-510-386-62	Sequence 62, Appl
43	46.5	14.4	586	1	US-10-131-826A-46	Sequence 46, Appl
44	46	14.2	616	1	US-10-982-545-5	Sequence 5, Appl
45	46	14.2	683	7	US-11-046-668-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-10-971-994-2
; Sequence 2, Application US/10971994
; Publication No. US20050250182A1
; GENERAL INFORMATION:
; APPLICANT: University of Michigan et al.
; APPLICANT: Kaufman, Randal
; APPLICANT: Kyungso, Lee
; APPLICANT: Mori, Kazutos
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, AND METHODS FOR
; MODULATING THE UNFOLDED PROTEIN RESPONSE
; FILE REFERENCE: UMV-2246
; CURRENT APPLICATION NUMBER: US/10/971,994
; CURRENT FILING DATE: 2004-10-21
; PRIOR APPLICATION NUMBER: PCT/2003/012640
; PRIOR FILING DATE: 2003-04-22
; PRIOR APPLICATION NUMBER: 60/375,098
; PRIOR FILING DATE: 2002-04-22
; PRIOR APPLICATION NUMBER: 60/374,880
; PRIOR FILING DATE: 2002-04-23
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-971-994-2

Query Match 28.1%; Score 91; DB 1; Length 376;
Best Local Similarity 40.3%; Pred. No. 0.00045;
Matches 27; Conservative 9; Mismatches 17; Indels 14; Gaps 4;

QY	11	GOELP--VP-----KTNLP-PRKRAK-TEDEKEQRRIERVLRNAAQTSRRKR	56
Db	33	GOALPLMVPAQRGASPEASGGLPQARKQRLTHLSPEEKALRRKLRVAAQTARDKK	92

QY	57	LEMEKLE 63
Db	93	ARMSELE 99

RESULT 2

US-10-131-826A-160
; Sequence 160, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen

```
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 160
; LENGTH: 605
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-131-826A-160

Query Match          21.5%; Score 69.5; DB 1; Length 605;
Best Local Similarity 34.4%; Pred. No. 0.18;
Matches 21; Conservative 8; Mismatches 15; Indels 17; Gaps 3;

QY 9 SNGQEL-----PVPKTNL-----PPRKRAKTEDEKEORRIERVLRNRAAAQTSRERKR 56
DB 134 SQAQLPNLPPLPKMNLVPPPHMPPREEEEEEERKEEVEK-----QEEEEEE 188

QY 57 L 57
DB 189 L 189

RESULT 3
US-11-169-013-2
; Sequence 2, Application US/11169013
; Publication No. US20050244971A1
; GENERAL INFORMATION:
; APPLICANT: Korea Kumho Petrochemical Co., Ltd.
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH ENHANCED STRESS TOLERANCE
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/11/169,013
; CURRENT FILING DATE: 2005-06-29
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Kopatentin 1.71
; SEQ ID NO 171
; LENGTH: 879
; TYPE: PRT
; ORGANISM: Simian Immunodeficiency Virus
US-11-022-562-340

Query Match          19.6%; Score 63.5; DB 7; Length 879;
Best Local Similarity 28.0%; Pred. No. 1.2;
Matches 21; Conservative 14; Mismatches 14; Indels 13; Gaps 3;

QY 2 KPAKKRKSNG-----QELPVPKTNLPPrKRAKTEDEKEORRIERVLRNRAA----AQ 49
DB 601 KDAQLNWGCAPRQVCHTTPVPNTSLTPKWDNETWQWE-RKVDFLSENTALPEEAQ 659
```

```
; SEQ ID NO 2
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-11-169-013-2

Query Match          20.2%; Score 65.5; DB 7; Length 431;
Best Local Similarity 38.6%; Pred. No. 0.34;
Matches 22; Conservative 10; Mismatches 16; Indels 9; Gaps 3;

QY 15 PVPKTNLPPrKRAKTEDEKE-ORRIERVLRNRAAAQTSRERKR-----LEMEKLE 63
DB 332 PVPYV-LNRGRSNTGLEKVIERRQRMKNRESAARSARKQAYTLELEAEIEKLK 387

RESULT 4
US-11-169-013-1
; Sequence 1, Application US/11169013
; Publication No. US20050244971A1
; GENERAL INFORMATION:
; APPLICANT: Korea Kumho Petrochemical Co., Ltd.
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH ENHANCED STRESS TOLERANCE
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/11/169,013
; CURRENT FILING DATE: 2005-06-29
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Kopatentin 1.71
; SEQ ID NO 1
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-11-169-013-1

Query Match          19.8%; Score 64; DB 7; Length 454;
Best Local Similarity 34.0%; Pred. No. 0.52;
Matches 17; Conservative 11; Mismatches 22; Indels 0; Gaps 0;

QY 15 PVPKTNLPPrKRAKTEDEKEORRIERVLRNRAAAQTSRERKRLEMEKLES 64
DB 353 PVPYMFGRVKTGAVLEKVIERRQRMKNRESAARSARKQAYTMELEA 402
```

```
RESULT 5
US-11-022-562-340
; Sequence 340, Application US/11022562
; Publication No. US20050249742A1
; GENERAL INFORMATION:
; APPLICANT: Ruprecht, Ruth M.
; APPLICANT: Shisong, Jiang
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; TITLE OF INVENTION: A CYTOTOXIC T LYMPHOCYTE IMMUNE RESPONSE
; FILE REFERENCE: DFN-043CN
; CURRENT APPLICATION NUMBER: US/11/022,562
; CURRENT FILING DATE: 2004-12-22
; PRIOR APPLICATION NUMBER: PCT/US03/20322
; PRIOR FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392718
; PRIOR FILING DATE: 2002-06-27
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 340
; LENGTH: 879
; TYPE: PRT
; ORGANISM: Simian Immunodeficiency Virus
US-11-022-562-340
```

```
Query Match          19.6%; Score 63.5; DB 7; Length 879;
Best Local Similarity 28.0%; Pred. No. 1.2;
Matches 21; Conservative 14; Mismatches 14; Indels 13; Gaps 3;

QY 2 KPAKKRKSNG-----QELPVPKTNLPPrKRAKTEDEKEORRIERVLRNRAA----AQ 49
DB 601 KDAQLNWGCAPRQVCHTTPVPNTSLTPKWDNETWQWE-RKVDFLSENTALPEEAQ 659
```

```
QY 50 TSRRKKLEMEKLES 64
Db 660 IQQEKMYELQKLS 674

RESULT 6
US-11-018-018-7
; Sequence 7, Application US/11018018
; Publication No. US20050250794A1
; GENERAL INFORMATION:
; APPLICANT: Napper, Andrew
; APPLICANT: Distefano, Peter
; APPLICANT: Curtis, Rory
; APPLICANT: Hixon, Jeffrey
; APPLICANT: McDonagh, Thomas
; TITLE OF INVENTION: METHODS OF TREATING A DISORDER
; FILE REFERENCE: 13407-052001
; CURRENT APPLICATION NUMBER: US/11/018,018
; CURRENT FILING DATE: 2004-12-20
; PRIOR APPLICATION NUMBER: US 60/530,945
; PRIOR FILING DATE: 2003-12-19
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-018-018-7

Query Match 18.2%; Score 59; DB 7; Length 400;
Best Local Similarity 37.5%; Pred. No. 1.6;
Matches 15; Conservative 11; Mismatches 12; Indels 2; Gaps 2;

QY 25 KRAKTEDEKQ--RRIERVLNRNAAQTSRRERKLEMEKLE 63
Db 18 RRLREEQRRRLRQVSRILR-KAAARSABEGRLLAESAD 56

RESULT 7
US-10-329-472-2
; Sequence 2, Application US/10329472
; Publication No. US20050244888A1
; GENERAL INFORMATION:
; APPLICANT: Li et al.
; TITLE OF INVENTION: Human Chemokine Beta-9
; FILE REFERENCE: PF131P2D1
; CURRENT APPLICATION NUMBER: US/10/329,472
; CURRENT FILING DATE: 2002-12-27
; PRIOR APPLICATION NUMBER: US 08/793,381
; PRIOR FILING DATE: 1997-05-19
; PRIOR FILING DATE: 1997-05-19
; PRIOR FILING DATE: 1996-02-29
; PRIOR APPLICATION NUMBER: US 08/294,251
; PRIOR FILING DATE: 1994-08-23
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 134
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-329-472-2

Query Match 17.3%; Score 56; DB 1; Length 134;
Best Local Similarity 25.3%; Pred. No. 1;
Matches 20; Conservative 15; Mismatches 20; Indels 24; Gaps 3;

QY 1 KKPAAKRRKSWGQELP-----VPKTNLPKRAKTE---DEKE-----QR 36
Db 39 KIPAKVVRYSRKQPSLGCISIPAILFLPKRSQAEKLPKELWVQMLQHLDTSPQK 98

QY 37 RIERVLNRNAAQTSRRERK 55
Db 37 RIERVLNRNAAQTSRRERK 55

Db 99 PAQCCKDRGASKTKKKGK 117

RESULT 8
US-11-074-176-82
; Sequence 82, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Kleenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Peril, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 82
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus
US-11-074-176-82

Query Match 17.1%; Score 55.5; DB 7; Length 194;
Best Local Similarity 28.4%; Pred. No. 1.8;
Matches 19; Conservative 10; Mismatches 35; Indels 3; Gaps 1;

QY 1 KKPAAKRRKSWGQELPVKTNLPKRAKTEDEKEQRRIERVLNRNAAQTSRRER---KRL 57
Db 22 KXAVKEAAKGEEKTKNNENQKLAETADLKEKNKQLEDKYLRSEAEIQNMQNRVTKERA 81

QY 58 EMEKLES 64
Db 82 QLIKYES 88

RESULT 9
US-10-510-386-6
; Sequence 6, Application US/10510386
; Publication No. US20050244922A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Clausen, Ib Groth
; APPLICANT: Jorgensen, Steen Troels
; APPLICANT: Olsen, Peter Bjarke
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: Improved Bacillus Host Cell
; FILE REFERENCE: 10294.204-US
; CURRENT APPLICATION NUMBER: US/10/510,386
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 6
; LENGTH: 625
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-10-510-386-6

Query Match 17.0%; Score 55; DB 1; Length 625;
Best Local Similarity 44.4%; Pred. No. 7.3;
Matches 12; Conservative 9; Mismatches 4; Indels 2; Gaps 1;

QY 25 KRAKTEDEKEQRRIERVLNRNAAQTS 51
Db 295 KRAKTKEEK--KKIEKELSNRISLTT 319

RESULT 10
```



```
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 2
; LENGTH: 188
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-131-826A-2

Query Match 16.4%; Score 53; DB 1; Length 188;
Best Local Similarity 28.6%; Pred. No. 3.2;
Matches 12; Conservative 13; Mismatches 15; Indels 2; Gaps 1;

Qy 25 KRAKTEDEKEQRIERVLNRRAAAQTSRRKRLM--EKLES 64
Db 37 KKLKAEVQSKKLEKKKTTITTESAGROQKKKIERQEEKLN 78

RESULT 14
US-11-022-562-217
; Sequence 217, Application US/11022562
; Publication No. US20050249742A1
; GENERAL INFORMATION:
; APPLICANT: Ruprecht, Ruth M.
; APPLICANT: Shisong, Jiang
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; FILE REFERENCE: DFN-043CN
; CURRENT APPLICATION NUMBER: US/11/022,562
; CURRENT FILING DATE: 2004-12-22
; PRIOR APPLICATION NUMBER: PCT/US03/20322
; PRIOR FILING DATE: 2003-06-27

; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 90 Human Secreted Proteins
; FILE REFERENCE: P2013P2C1
; CURRENT APPLICATION NUMBER: US/10/986,501
; CURRENT FILING DATE: 2004-11-12
; PRIOR APPLICATION NUMBER: US/10/621,363
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: 09/969,730
; PRIOR FILING DATE: 2001-10-06
; PRIOR APPLICATION NUMBER: 09/774,639
; PRIOR FILING DATE: 2001-02-01
; PRIOR APPLICATION NUMBER: 60/238,291
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 09/244,112
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: PCT/US98/16235
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/056,371
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,732
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,366
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,364
; PRIOR FILING DATE: 1997-08-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 373
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 124
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-986-501-124

Query Match 15.7%; Score 51; DB 1; Length 312;
Best Local Similarity 25.5%; Pred. No. 9.3;
Matches 12; Conservative 11; Mismatches 22; Indels 2; Gaps 1;

Qy 13 ELPPVPTNL--PPRKAKTEDEKEQRIERVLNRRAAAQTSRRKRL 57
Db 233 KLPEFSASLPNPPSKKIKLSDEPVEAKEDYTKFNTKDLKTEKNSKM 279

Search completed: November 22, 2005, 15:54:45
Job time : 4 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2005, 15:38:17 ; Search time 111 Seconds
(without alignments)
240.911 Million cell updates/sec

Title: US-10-663-450-6_COPY_53_116

Perfect score: 324

Sequence: 1 KKPAAKRSWGQELPVKTN.....RAAAQTSRRKRLEMEKLES 64

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries.

Database :

1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pap:*
2: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pap:*
3: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pap:*
4: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pap:*
5: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pap:*
6: /cgn2_6/prodata/1/pubpaa/US11_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	324	100.0	349	3	US-09-816-277-4
2	324	100.0	349	3	US-09-816-277-6
3	324	100.0	349	4	US-10-663-450-4
4	324	100.0	349	4	US-10-663-450-6
5	317	97.8	342	3	US-09-816-277-19
6	317	97.8	342	4	US-10-663-450-19
7	317	97.8	342	3	US-09-816-277-16
8	317	97.8	342	4	US-10-663-450-16
9	283	87.3	451	3	US-09-816-277-2
10	283	87.3	451	3	US-09-816-277-5
11	283	87.3	451	4	US-10-663-450-2
12	283	87.3	451	4	US-10-663-450-5
13	277	85.5	409	3	US-09-533-029-104
14	277	85.5	409	4	US-10-295-403-158
15	277	85.5	409	4	US-10-412-699B-684
16	250	77.2	174	4	US-10-767-701-61238
17	171.5	52.9	200	4	US-10-369-493-21867
18	170	52.5	68	3	US-09-816-277-60
19	170	52.5	68	4	US-10-663-450-60
20	117	36.1	203	4	US-10-437-963-151695
21	117	36.1	203	5	US-10-732-923-13694
22	117	36.1	203	5	US-10-732-923-13695
23	116	35.8	168	3	US-09-334-455-102
24	116	35.8	168	4	US-10-325-066A-206
25	116	35.8	168	4	US-10-374-780A-2356
26	116	35.8	168	5	US-10-732-923-13692
27	116	35.8	168	5	US-10-225-066A-206

28 111.5 34.4 192 3 US-09-772-656-2 Sequence 2, Appli
29 111.5 34.4 192 3 US-09-772-656-6 Sequence 6, Appli
30 111.5 34.4 192 3 US-09-772-656-10 Sequence 10, Appli
31 111.5 34.4 192 4 US-10-396-199A-2 Sequence 2, Appli
32 111.5 34.4 192 4 US-10-396-199A-6 Sequence 6, Appli
33 111.5 34.4 192 4 US-10-396-199A-10 Sequence 10, Appli
34 111 34.3 147 4 US-10-425-115-187302 Sequence 187302,
35 109 33.6 109 5 US-10-732-923-13697 Sequence 13697, A
36 109 33.6 196 4 US-10-425-115-197786 Sequence 197786,
37 108 33.3 176 4 US-10-437-963-164243 Sequence 164243,
38 106.5 32.9 170 4 US-10-310-154-565 Sequence 565, App
39 104.5 32.3 191 4 US-10-424-599-177263 Sequence 177263,
40 104 32.1 181 5 US-10-732-923-450 Sequence 450, App
41 101 31.2 217 4 US-10-425-115-197783 Sequence 197783,
42 100 30.9 143 5 US-10-856-499-1930 Sequence 1930, Ap
43 99 30.6 114 4 US-10-425-115-246430 Sequence 246430,
44 99 30.6 360 4 US-10-437-963-165206 Sequence 165206,
45 98 30.2 149 5 US-10-732-923-13693 Sequence 13693, A

ALIGNMENTS

RESULT 1

US-09-816-277-4
; Sequence 4, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Salohelmo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-816-277-4

Query Match 100.0%; Score 324; DB 3; Length 349;
Best Local Similarity 100.0%; Pred. No. 5.1e-25;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKPAAKRSWGQELPVKTNLPKRKAKTEDEKQRIERVLNRAAAQTSRRKRLEME 60
DB 53 KKPAAKRSWGQELPVKTNLPKRKAKTEDEKQRIERVLNRAAAQTSRRKRLEME 112
QY 61 KLES 64
DB 113 KLES 116

RESULT 2

US-09-816-277-6
; Sequence 6, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Salohelmo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE OF INVENTION: Proteins by Recombinant Eukaryotic Cells

```
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-816-277-6

Query Match      100.0%; Score 324; DB 3; Length 349;
Best Local Similarity 100.0%; Pred. No. 5.1e-25;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKPAAKRRKSWGQELPVPKTNLPKRKAKTEDEKEQRRRIERVLRNRAAAQTSRKRKLEME 60
Db 53 KKPAAKRRKSWGQELPVPKTNLPKRKAKTEDEKEQRRRIERVLRNRAAAQTSRKRKLEME 112
QY 61 KLES 64
Db 113 KLES 116

RESULT 3
US-10-663-450-4
; Sequence 4, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-663-450-4

Query Match      100.0%; Score 324; DB 4; Length 349;
Best Local Similarity 100.0%; Pred. No. 5.1e-25;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKPAAKRRKSWGQELPVPKTNLPKRKAKTEDEKEQRRRIERVLRNRAAAQTSRKRKLEME 60
Db 53 KKPAAKRRKSWGQELPVPKTNLPKRKAKTEDEKEQRRRIERVLRNRAAAQTSRKRKLEME 112
QY 61 KLES 64
Db 113 KLES 116

RESULT 4
US-10-663-450-6
; Sequence 6, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
```

```
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-663-450-6

Query Match      100.0%; Score 324; DB 4; Length 349;
Best Local Similarity 100.0%; Pred. No. 5.1e-25;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKPAAKRRKSWGQELPVPKTNLPKRKAKTEDEKEQRRRIERVLRNRAAAQTSRKRKLEME 60
Db 53 KKPAAKRRKSWGQELPVPKTNLPKRKAKTEDEKEQRRRIERVLRNRAAAQTSRKRKLEME 112
QY 61 KLES 64
Db 113 KLES 116

RESULT 5
US-09-816-277-19
; Sequence 19, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-09-816-277-19

Query Match      97.8%; Score 317; DB 3; Length 342;
Best Local Similarity 96.9%; Pred. No. 2.6e-24;
Matches 62; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKPAAKRRKSWGQELPVPKTNLPKRKAKTEDEKEQRRRIERVLRNRAAAQTSRKRKLEME 60
Db 47 KKPAAKRRKSWGQELPVPKTNLPKRKAKTEDEKEQRRRIERVLRNRAAAQTSRKRKLEME 106
QY 61 KLES 64
Db 107 KLEN 110

RESULT 6
US-10-663-450-19
; Sequence 19, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
```

Publication No.	US20040186070A1
GENERAL INFORMATION:	
APPLICANT: Penttila, Merja E.	
APPLICANT: Ward, Michael	
APPLICANT: Wang, Huaming	
APPLICANT: Valkonen, Mari J.	
APPLICANT: Saloheimo, Markku	
TITLE OF INVENTION: Increased Production of Secreted	
PROTEINS BY RECOMBINANT EUKARYOTIC CELLS	
CURRENT APPLICATION NUMBER: US/10/663,450	
PRIOR FILING DATE: 2003-09-15	
PRIOR APPLICATION NUMBER: US 09/534,692	
NUMBER OF SEQ ID NOS: 63	
SOFTWARE: FastSeq for Windows Version 4.0	
SEQ ID NO 19	
LENGTH: 342	
TYPE: PRT	
ORGANISM: Aspergillus niger	
US-10-663-450-19	
Query Match	97.8%; Score 317; DB 4; Length 342;
Best Local Similarity	96.9%; Pred. No. 2.6e-24;
Matches	62; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY	1 KKPAAKRSWGQELPVPKTNLPKRKAKTEDEKQRIERVLRNRAAAQTSRERKRLME 60
DB	47 KKPAAKRSWGQELPVPKTNLPKRKAKTEDEKQRIERVLRNRAAAQTSRERKRLME 106
QY	61 KLES 64
DB	107 KLEN 110
RESULT 7	
US-09-816-277-16	
Sequence 16, Application US/09816277	
Publication No. US20010034045A1	
GENERAL INFORMATION:	
APPLICANT: Penttila, Merja E.	
APPLICANT: Ward, Michael	
APPLICANT: Wang, Huaming	
APPLICANT: Valkonen, Mari J.	
APPLICANT: Saloheimo, Markku	
TITLE OF INVENTION: Increased Production of Secreted	
PROTEINS BY RECOMBINANT EUKARYOTIC CELLS	
CURRENT APPLICATION NUMBER: US/09/816,277	
PRIOR FILING DATE: 2001-03-23	
PRIOR APPLICATION NUMBER: US 09/534,692	
NUMBER OF SEQ ID NOS: 63	
SOFTWARE: FastSeq for Windows Version 4.0	
SEQ ID NO 16	
LENGTH: 386	
TYPE: PRT	
ORGANISM: Aspergillus niger	
US-09-816-277-16	
Query Match	97.8%; Score 317; DB 3; Length 386;
Best Local Similarity	96.9%; Pred. No. 2.9e-24;
Matches	62; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY	1 KKPAAKRSWGQELPVPKTNLPKRKAKTEDEKQRIERVLRNRAAAQTSRERKRLME 60
DB	91 KKPAAKRSWGQELPVPKTNLPKRKAKTEDEKQRIERVLRNRAAAQTSRERKRLME 150
QY	61 KLES 64
DB	151 KLEN 154
RESULT 8	
US-10-663-450-16	
Sequence 16, Application US/10663450	
Publication No. US20040186070A1	
GENERAL INFORMATION:	
APPLICANT: Penttila, Merja E.	
APPLICANT: Ward, Michael	
APPLICANT: Wang, Huaming	
APPLICANT: Valkonen, Mari J.	
APPLICANT: Saloheimo, Markku	
TITLE OF INVENTION: Increased Production of Secreted	
PROTEINS BY RECOMBINANT EUKARYOTIC CELLS	
CURRENT APPLICATION NUMBER: US/10/663,450	
PRIOR FILING DATE: 2003-09-15	
PRIOR APPLICATION NUMBER: US 09/534,692	
NUMBER OF SEQ ID NOS: 63	
SOFTWARE: FastSeq for Windows Version 4.0	
SEQ ID NO 19	
LENGTH: 342	
TYPE: PRT	
ORGANISM: Aspergillus niger	
US-10-663-450-19	
Query Match	97.8%; Score 317; DB 3; Length 386;
Best Local Similarity	96.9%; Pred. No. 2.9e-24;
Matches	62; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY	1 KKPAAKRSWGQELPVPKTNLPKRKAKTEDEKQRIERVLRNRAAAQTSRERKRLME 60
DB	91 KKPAAKRSWGQELPVPKTNLPKRKAKTEDEKQRIERVLRNRAAAQTSRERKRLME 150
QY	61 KLES 64
DB	151 KLEN 154
RESULT 9	
US-09-816-277-2	
Sequence 2, Application US/09816277	
Publication No. US20010034045A1	
GENERAL INFORMATION:	
APPLICANT: Penttila, Merja E.	
APPLICANT: Ward, Michael	
APPLICANT: Wang, Huaming	
APPLICANT: Valkonen, Mari J.	
APPLICANT: Saloheimo, Markku	
TITLE OF INVENTION: Increased Production of Secreted	
PROTEINS BY RECOMBINANT EUKARYOTIC CELLS	
CURRENT APPLICATION NUMBER: US/09/816,277	
PRIOR FILING DATE: 2001-03-23	
PRIOR APPLICATION NUMBER: US 09/534,692	
NUMBER OF SEQ ID NOS: 63	
SOFTWARE: FastSeq for Windows Version 4.0	
SEQ ID NO 2	
LENGTH: 451	
TYPE: PRT	
ORGANISM: Trichoderma reesei	
US-09-816-277-2	
Query Match	87.3%; Score 283; DB 3; Length 451;
Best Local Similarity	87.3%; Pred. No. 1e-20;
Matches	55; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY	1 KKPAAKRSWGQELPVPKTNLPKRKAKTEDEKQRIERVLRNRAAAQTSRERKRLME 60
DB	85 KKPAAKRSWGQELPVPKTNLPKRKAKTEDEKQRIERVLRNRAAAQTSRERKRLME 144
QY	61 KLE 63
DB	145 ALE 147

```

RESULT 10
US-09-816-277-5
; Sequence 5, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-09-816-277-5

Query Match      87.3%; Score 283; DB 3; Length 451;
Best Local Similarity 87.3%; Pred. No. 1e-20;
Matches 55; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 KKPAPKRSWGQELPVPTKNLPPEKRAKTEDEKEQRIERVLRNRAAAQTSRRKRLME 60
Db 85 KKPVPKRSWGQVLPPEPTKNLPPEKRAKTEDEKEQRRVLRNRAAQSSRRKRLVE 144

```

```

RESULT 11
US-10-663-450-2
; Sequence 2, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
; US-10-663-450-2

Query Match      87.3%; Score 283; DB 4; Length 451;
Best Local Similarity 87.3%; Pred. No. 1e-20;
Matches 55; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy      1 KKPAPKRSKSGQELPVPKTNLP PPKRAKTEDEKESQRIERVLNRRAAAQTSRRKRLME 60
Db      85 KKPVKRKSGQVLPEPKTNLP PPKRAKTEDEKESQRRVERVLNRRAAQSSRRKRLVE 144
Qy      61 KLE 63
      ==

```

Db 145 ALE 147

RESULT 12

US-10-663-450-5
; Sequence 5, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttilä, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; ; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-663-450-5

Query Match 87.3%; Score 283; DB 4; Length 451;
Best Local Similarity 87.3%; Pred.No.1e-20; Indels 0; Gaps 0;
Matches 55; Conservative 3; Mismatches 5;

QY 1 KKPAKKRSKGQELVPKTNLPPRKRAKTDEKEQRRIERVLNRRAAAQTSRERKRLEME 60
||| |||||||
DB 85 KKPVKRRSKGQVLPKTNLPPRKRAKTDEKEQRRIERVLNRRAAAQSRRERKRLEVE 144
||| |||||||

QY 61 KLE 63
||
DB 145 ALE 147

RESULT 13

US-09-533-029-104
; Sequence 104, Application US/09533029
; Publication No. US20030046723A1
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Broun, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Keddie, James
; APPLICANT: Pineda, Omaira
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Reuber, Lynne
; TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES
; FILE REFERENCE: MBI-010
; CURRENT APPLICATION NUMBER: US/09/533,029
; CURRENT FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 60/125,814
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 104
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1034

THIS PAGE BLANK (USPTO)

Result No.	Score	Query No.	Length	DB	ID	Description	
1	77	23.8	376	1	US-10-571-994-2	Sequence 2, Appli	
2	66.5	20.6	605	1	US-10-131-826A-160	Sequence 160, App	
3	65	20.1	431	7	US-11-169-013-2	Sequence 2, Appli	
4	62.5	19.3	454	7	US-11-169-013-1	Sequence 1, Appli	
5	59	18.3	747	7	US-11-182-592-2	Sequence 2, Appli	
6	58.5	18.1	903	1	US-10-689-742-142	Sequence 142, App	
7	57	17.6	312	1	US-10-986-501-124	Sequence 124, App	
8	57	17.6	400	7	US-11-018-018-7	Sequence 7, Appli	
9	56.5	17.5	180	1	US-10-689-742-84	Sequence 84, Appli	
10	56.5	17.5	1618	1	US-10-984-645-2	Sequence 2, Appli	
11	54	16.7	21	1	US-10-995-951A-33	Sequence 33, Appli	
12	54	16.7	170	1	US-10-689-742-16	Sequence 16, Appli	
13	53.5	16.6	569	1	US-10-632-150-2	Sequence 2, Appli	
14	51.5	15.9	615	1	US-10-982-545-14	Sequence 14, Appli	
15	51.5	15.9	879	7	US-11-022-562-340	Sequence 340, App	
16	51	15.8	2897	7	US-10-499-715-2	Sequence 2, Appli	
17	50.5	15.6	296	7	US-11-078-280-2	Sequence 2, Appli	
18	50.5	15.6	327	7	US-11-090-878-16	Sequence 16, Appli	
19	50.5	15.6	874	7	US-11-012-762-8	Sequence 8, Appli	
20	50	15.5	790	1	US-10-131-826A-204	Sequence 204, App	
21	49.5	15.3	128	7	US-11-105-864-4	Sequence 4, Appli	
22	49	15.2	188	1	US-10-131-826A-2	Sequence 2, Appli	
23	49	15.2	194	7	US-11-074-176-82	Sequence 82, Appli	
24	49	15.2	625	1	US-10-510-386-6	Sequence 6, Appli	
25	49	15.2	1302	7	US-11-004-057-6	Sequence 6, Appli	

APPLICANT: DeForge,Laura
APPLICANT: Desnoyers,Luc
APPLICANT: Filvaroff,Ellen
APPLICANT: Gao,Wei-Qiang
APPLICANT: Gerritsen,Mary E.
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Sherwood,Steven
APPLICANT: Smith,Victoria
APPLICANT: Stewart,Timothy A.
APPLICANT: Tumas,Daniel
APPLICANT: Watanabe,Colin K
APPLICANT: Wood,William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C128
CURRENT APPLICATION NUMBER: US/10/131,826A
CURRENT FILING DATE: 2002-04-24
PRIOR APPLICATION NUMBER: 60/049911
PRIOR FILING DATE: 1997-06-18
PRIOR APPLICATION NUMBER: 60/056974
PRIOR FILING DATE: 1997-08-26
PRIOR APPLICATION NUMBER: 60/059113
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059115
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059117
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059122
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059184
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059352
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 60/059588
PRIOR FILING DATE: 1997-09-19
Remaining prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 160
LENGTH: 605
TYPE: PRT
ORGANISM: Homo Sapien
US-10-131-826A-160
Query Match 20.6%; Score 66.5; DB 1; Length 605;
Best Local Similarity 34.1%; Pred. No. 0.41;
Matches 15; Conservative 9; Mismatches 15; Indels 5; Gaps 1;
QY 15 LPEPTNLPKRAKTEDEKEQRRVRLNRRRAAQSRRKRL 58
Db 151 LVEPPWHMPRBEEREEEREKEVEK-----QEEEEEEL 189
RESULT 3
US-11-169-013-2
Sequence 2, Application US/11169013
Publication No. US20050244971A1
GENERAL INFORMATION:
APPLICANT: Korea Kumho Petrochemical Co., Ltd.
TITLE OF INVENTION: TRANSGENIC PLANTS WITH ENHANCED STRESS TOLERANCE
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/11/169,013
CURRENT FILING DATE: 2005-06-29
NUMBER OF SEQ ID NOS: 34
SOFTWARE: Kopatentin 1.71
SEQ ID NO 2
LENGTH: 431
TYPE: PRT
ORGANISM: Arabidopsis thaliana

US-11-169-013-2
Query Match 20.1%; Score 65; DB 7; Length 431;
Best Local Similarity 43.9%; Pred. No. 0.41;
Matches 18; Conservative 9; Mismatches 10; Indels 4; Gaps 2;
QY 26 KRAKTEDEKE-QRRVRLNRRRAAQSRRKX---LEVEA 62
Db 341 RRSNTGLEKVIERRQRRMIKNRESAARSRAKQAYTLEEA 381
RESULT 4
US-11-169-013-1
Sequence 1, Application US/11169013
Publication No. US20050244971A1
GENERAL INFORMATION:
APPLICANT: Korea Kumho Petrochemical Co., Ltd.
TITLE OF INVENTION: TRANSGENIC PLANTS WITH ENHANCED STRESS TOLERANCE
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/11/169,013
CURRENT FILING DATE: 2005-06-29
NUMBER OF SEQ ID NOS: 34
SOFTWARE: Kopatentin 1.71
SEQ ID NO 1
LENGTH: 454
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-11-169-013-1
Query Match 19.3%; Score 62.5; DB 7; Length 454;
Best Local Similarity 36.6%; Pred. No. 0.82;
Matches 15; Conservative 11; Mismatches 12; Indels 3; Gaps 1;
QY 25 RKRAKTEDEKEQRRVRLNRRRAAQSRRKX---LEVEA 62
Db 362 RKTGAVLEKVIERRQRRMIKNRESAARSRAKQAYTMELEA 402
RESULT 5
US-11-182-592-2
Sequence 2, Application US/11182592
Publication No. US20050250153A1
GENERAL INFORMATION:
APPLICANT: Nezu, Jun-ichi
APPLICANT: Ose, Asuka
TITLE OF INVENTION: NOVEL FETAL GENES
FILE REFERENCE: 14875-089001 (previously 06501-0890001)
CURRENT APPLICATION NUMBER: US/11/182,592
CURRENT FILING DATE: 2005-07-14
PRIOR APPLICATION NUMBER: US/09/974,143
PRIOR FILING DATE: 2001-10-09
PRIOR APPLICATION NUMBER: PCT/JP00/02281
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 11/103356
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 36
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 747
TYPE: PRT
ORGANISM: Homo sapiens
US-11-182-592-2
Query Match 18.3%; Score 59; DB 7; Length 747;
Best Local Similarity 32.8%; Pred. No. 3.5;
Matches 22; Conservative 11; Mismatches 28; Indels 6; Gaps 3;
QY 1 EKKPVKKRSGVLPPEKTNLPKRAKTEDEKEQRRVRLNRRRAAQSRR---ERK 56
Db 643 EKKSVAEGLS-GSLVQEP-FQLATEKRAKERQLEKRMALVEAQAQQLAEARLQEEOK 700
QY 57 RLEVEAL 63
: | : |


```

Db          701 KEELARL 707

RESULT 6
US-10-689-742-142
; Sequence 142, Application US/10689742
; Publication No. US20050250180A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M
; APPLICANT: LaVallie, Edward R
; APPLICANT: Racie, Lisa A
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SCRPTED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 00766.000091.10
; CURRENT APPLICATION NUMBER: US/10/689,742
; CURRENT FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: 09/746,783
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 142
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-689-742-142

Query Match      18.1%; Score 58.5; DB 1; Length 903;
Best Local Similarity 30.5%; Pred. No. 4.9;
Matches 18; Conservative 8; Mismatches 26; Indels 7; Gaps 1;

QY      6 KKRKSGVGLPEPTNLPPEKRAETEKEQRRVRLNRPRAQSSPERKLEVALE 64
DB      51 KKR-----PEPVIQKGNNGRTDLKQOSTRESVMVSPKRGLSSEKDNIRQAI 102

RESULT 7
US-10-986-501-124
; Sequence 124, Application US/10986501
; Publication No. US20050244845A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 90 Human Secreted Proteins
; FILE REFERENCE: P2013P2C1
; CURRENT APPLICATION NUMBER: US/10/986,501
; CURRENT FILING DATE: 2004-11-12
; PRIOR APPLICATION NUMBER: US/10/621,363
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: 09/969,730
; PRIOR FILING DATE: 2001-10-06
; PRIOR APPLICATION NUMBER: 09/774,639
; PRIOR FILING DATE: 2001-02-01
; PRIOR APPLICATION NUMBER: 60/238,291
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 09/244,112
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: PCT/US98/16235
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/056,371
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,732
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,366
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,364
; PRIOR FILING DATE: 1997-08-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 373
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 124

```

```
Query Match 17.5%; Score 56.5; DB 1; Length 180;
Best Local Similarity 29.1%; Pred. No. 1.3;
Matches 23; Conservative 12; Mismatches 27; Indels 17; Gaps 4;

QY 1 EKKPVKKR-----KSWGQVLP-PKTNLPKRKRAKTEDEKEQRRVRLNRRAA 49
DB 41 EVKQDKRASQGSFEVLKSPDLSPESWMLSPPKK-DTSLEELQKLEAAEERRKTQ 99

QY 50 QSS-----RERKLEVEAL 63
DB 100 EAQVLQLAERREHEREVL 118

RESULT 10
US-10-984-645-2
; Sequence 2, Application US/10984645
; Publication No. US20050244386A1
; GENERAL INFORMATION:
; APPLICANT: Habener, Joel
; APPLICANT: Zulewski, Hendrik
; APPLICANT: Abraham, Elizabeth
; APPLICANT: Vallejo, Mario
; TITLE OF INVENTION: METHOD OF TRANSPLANTING IN A MAMMAL AND TREATING DIABETES MELLITU
; TITLE OF INVENTION: BY ADMINISTERING A PSEUDO-ISLET LIKE AGGREGATE DIFFERENTIATED FR
; TITLE OF INVENTION: A NESTIN-POSITIVE PANCREATIC STEM CELL
; FILE REFERENCE: 3284/1223
; CURRENT APPLICATION NUMBER: US/10/984,645
; CURRENT FILING DATE: 2004-11-09
; PRIOR APPLICATION NUMBER: US 09/731,255
; PRIOR FILING DATE: 2000-12-06
; PRIOR APPLICATION NUMBER: US 60/169,082
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: US 60/215,109
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: US 60/239,880
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 1618
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-984-645-2

Query Match 17.5%; Score 56.5; DB 1; Length 1618;
Best Local Similarity 29.7%; Pred. No. 16;
Matches 22; Conservative 14; Mismatches 23; Indels 15; Gaps 4;

QY 1 EKKPVKKRKSQGO-----VLPEKTNLPKRKRAKTEDEKEQRRVRLNRRAAQSSRR 55
DB 758 EKETQQRRLSGDQDQMTLRPPEKVDLEP---LKSIDQEIARPLEN--ENQEFKLSKEE 812

QY 56 -----KRLEVEALE 64
DB 813 SVEAVKSLTEILE 826

RESULT 11
US-10-995-951A-33
; Sequence 33, Application US/10995951A
; Publication No. US20050245732A1
; GENERAL INFORMATION:
; APPLICANT: Hamoufa, A. et al.
; TITLE OF INVENTION: A Repressor-Mediated Regulation System for Control of Gene Expres
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 1096.021B
; CURRENT APPLICATION NUMBER: US/10/995,951A
; CURRENT FILING DATE: 2004-11-23
; PRIOR APPLICATION NUMBER: PCT/CA02/01807
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: PCT/CA02/00740
; PRIOR FILING DATE: 2002-05-23
```

```
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 33
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Nicotiana sp.
US-10-995-951A-33

Query Match 16.7%; Score 54; DB 1; Length 21;
Best Local Similarity 58.8%; Pred. No. 0.21;
Matches 10; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 41 RVLNRRAAQSSRRKR 57
DB 5 RLVNRRESAQLSRQKK 21

RESULT 12
US-10-689-742-16
; Sequence 16, Application US/10689742
; Publication No. US20050250180A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M
; APPLICANT: Lavallie, Edward R
; APPLICANT: Racie, Lisa A
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 00766.000091.10
; CURRENT APPLICATION NUMBER: US/10/689,742
; CURRENT FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: 09/746,783
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-689-742-16

Query Match 16.7%; Score 54; DB 1; Length 170;
Best Local Similarity 35.7%; Pred. No. 2.3;
Matches 15; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 15 LPEPKTNLPKRKRAKTEDEKEQRRVRLNRRAAQSSRRK 56
DB 14 LQESQTLEHTKGALTEQHERVHRLTEHVNAMRGLQSSKELK 55

RESULT 13
US-10-632-150-2
; Sequence 2, Application US/10632150
; Publication No. US20050251871A1
; GENERAL INFORMATION:
; APPLICANT: Chiaur, D.
; APPLICANT: Pagano, M.
; APPLICANT: Latres, E.
; TITLE OF INVENTION: NOVEL UBIQUITIN LIGASES AS THERAPEUTIC TARGETS
; FILE REFERENCE: 5914-081
; CURRENT APPLICATION NUMBER: US/10/632,150
; CURRENT FILING DATE: 2003-07-30
; PRIOR APPLICATION NUMBER: US/09/385,219
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: 60/098,355
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: 60/118,568
; PRIOR FILING DATE: 1999-02-03
; PRIOR APPLICATION NUMBER: 60/124,449
; PRIOR FILING DATE: 1999-03-15
```

```
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-632-150-2

Query Match      16.6%; Score 53.5; DB 1; Length 569;
Best Local Similarity 28.8%; Pred. No. 10;
Matches 15; Conservative 8; Mismatches 26; Indels 3; Gaps 1;

QY 7 KKKSQGVLP---PEPKTNLPKRKRAKTEDEKQRRVERVLRRAAQSSRR 55
Db 208 ERKGQYLFKNKPPDGNAPPNSFYALYPKIIDTETIESNRCGRSLQR 259

RESULT 14
US-10-982-545-14
; Sequence 14, Application US/10982545
; Publication No. US20050244890A1
; GENERAL INFORMATION:
; APPLICANT: Davies, Huw Alun
; APPLICANT: McGuire, James
; APPLICANT: Simonsen, Anja Hviid
; APPLICANT: Blennow, Kaj
; APPLICANT: Podust, Vladimir
; APPLICANT: Ciphergen Biosystems, Inc.
; TITLE OF INVENTION: Biomarkers for Alzheimer's Disease
; FILE REFERENCE: 016866-011550US
; CURRENT APPLICATION NUMBER: US/10/982,545
; PRIOR FILING DATE: 2004-11-06
; PRIOR APPLICATION NUMBER: US 60/518,360
; PRIOR FILING DATE: 2003-11-07
; PRIOR APPLICATION NUMBER: US 60/526,753
; PRIOR FILING DATE: 2003-12-02
; PRIOR APPLICATION NUMBER: US 60/546,423
; PRIOR FILING DATE: 2004-02-19
; PRIOR APPLICATION NUMBER: US 60/547,250
; PRIOR FILING DATE: 2004-02-23
; PRIOR APPLICATION NUMBER: US 60/558,896
; PRIOR FILING DATE: 2004-04-02
; PRIOR APPLICATION NUMBER: US 60/572,617
; PRIOR FILING DATE: 2004-05-18
; PRIOR APPLICATION NUMBER: US 60/586,503
; PRIOR FILING DATE: 2004-07-08
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 615
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: nerve growth factor inducible Neurosecretory
; OTHER INFORMATION: protein vaccinia virus growth factor (VGF)
; OTHER INFORMATION: precursor
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(22)
; OTHER INFORMATION: signal peptide
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (23)..(615)
; OTHER INFORMATION: nerve growth factor inducible Neurosecretory
; OTHER INFORMATION: protein vaccinia virus growth factor (VGF)
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (373)..(417)
; OTHER INFORMATION: biomarker peptide 4808 Da, 4812 Da (Q10),
; OTHER INFORMATION: processed fragment of nerve growth factor
; OTHER INFORMATION: inducible Neurosecretory protein vaccinia virus
; OTHER INFORMATION: growth factor (VGF)
US-10-982-545-14

Query Match      15.9%; Score 51.5; DB 1; Length 615;
Best Local Similarity 31.5%; Pred. No. 19;
Matches 17; Conservative 8; Mismatches 14; Indels 15; Gaps 3;

QY 7 KKKSQGVLPPEPKTNLPKR---KRAKTEDEKEQRR-----VERVLRNR 46
Db 562 RRRHYHALP-PSRRYGRQAQRAQBEAEARRLQEQBELENYIEHVLR 614

RESULT 15
US-11-022-562-340
; Sequence 340, Application US/11022562
; Publication No. US20050249742A1
; GENERAL INFORMATION:
; APPLICANT: Ruprecht, Ruth M.
; APPLICANT: Shisong, Jiang
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; TITLE OF INVENTION: A CYTOTOXIC T LYMPHOCYTE IMMUNE RESPONSE
; FILE REFERENCE: DFN-043CN
; CURRENT APPLICATION NUMBER: US/11/022,562
; CURRENT FILING DATE: 2004-12-22
; PRIOR APPLICATION NUMBER: PCT/US03/20322
; PRIOR FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392718
; PRIOR FILING DATE: 2002-06-27
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 340
; LENGTH: 879
; TYPE: PRT
; ORGANISM: Simian Immunodeficiency Virus
US-11-022-562-340

Query Match      15.9%; Score 51.5; DB 7; Length 879;
Best Local Similarity 28.8%; Pred. No. 28;
Matches 19; Conservative 12; Mismatches 22; Indels 13; Gaps 3;

QY 10 SWG-----QVLPEPKTNLPKRKRAKTEDEKQRRVERVLRNRRA----AQSSRRERKR 57
Db 608 AWGCAFRQVCHTTPWPNTSLTPKWDNETQWE-RKVDLFLEENITALPEEAQIQQEKNM 666
QY 58 LEVEAL 63
Db 667 YELQKL 672

Search completed: November 22, 2005, 15:54:45
Job time : 5 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2005, 15:38:17 ; Search time 111 Seconds
(without alignments)
240.911 Million cell updates/sec

Title: us-10-663-450-5_COPY_84_147

Perfect score: 323

Sequence: 1 EKKPVKRSWGQVLPEPT.....NRRQAQSSRRKRLVEALE 64

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : 1: /cgn2_6/prodata/1/pubaa/US07_PUBCOMB.pap:*
2: /cgn2_6/prodata/1/pubaa/US08_PUBCOMB.pap:*
3: /cgn2_6/prodata/1/pubaa/US09_PUBCOMB.pap:*
4: /cgn2_6/prodata/1/pubaa/US10A_PUBCOMB.pap:*
5: /cgn2_6/prodata/1/pubaa/US10B_PUBCOMB.pap:*
6: /cgn2_6/prodata/1/pubaa/US11_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	323	100.0	451	3	US-09-816-277-2
2	323	100.0	451	3	US-09-816-277-5
3	323	100.0	451	4	US-10-663-450-2
4	323	100.0	451	4	US-10-663-450-5
5	307	95.0	409	3	US-09-533-029-104
6	307	95.0	409	4	US-10-295-403-158
7	307	95.0	409	4	US-10-412-699B-684
8	292	90.4	342	3	US-09-816-277-19
9	292	90.4	342	4	US-10-663-450-19
10	292	90.4	386	3	US-09-816-277-16
11	292	90.4	386	4	US-10-663-450-16
12	288	89.2	349	3	US-10-816-277-4
13	288	89.2	349	3	US-09-816-277-6
14	288	89.2	349	4	US-10-663-450-4
15	288	89.2	349	4	US-10-663-450-6
16	268	83.0	174	4	US-10-767-701-61238
17	175	54.2	68	3	US-09-816-277-60
18	175	54.2	68	4	US-10-663-450-60
19	175	54.2	200	4	US-10-369-493-21867
20	108	33.4	147	4	US-10-425-115-187302
21	108	33.4	192	3	US-09-772-656-2
22	108	33.4	192	3	US-09-772-656-6
23	108	33.4	192	3	US-09-772-656-10
24	108	33.4	192	4	US-10-396-199A-2
25	108	33.4	192	4	US-10-396-199A-6
26	107	33.1	203	4	US-10-437-963-151695
27	107	33.1	203	4	US-10-437-963-151695

28 107 33.1 203 5 US-10-732-923-13694 Sequence 13694, A
29 107 33.1 203 5 US-10-732-923-13695 Sequence 13695, A
30 106 32.8 168 3 US-09-934-453-102 Sequence 102, App
31 106 32.8 168 4 US-10-225-066A-206 Sequence 206, App
32 106 32.8 168 4 US-10-374-780A-2356 Sequence 2356, App
33 106 32.8 168 5 US-10-732-923-13692 Sequence 13692, A
34 106 32.8 168 5 US-10-225-066A-206 Sequence 206, App
35 103.5 32.0 191 4 US-10-424-599-177263 Sequence 177263, App
36 103 31.9 170 4 US-10-310-154-565 Sequence 565, App
37 102 31.6 109 5 US-10-732-923-13697 Sequence 13697, A
38 102 31.6 196 4 US-10-425-115-197786 Sequence 197786, App
39 101 31.3 665 4 US-10-425-115-190929 Sequence 190929, A
40 101 31.3 672 4 US-10-425-114-58469 Sequence 58469, A
41 99 30.7 143 5 US-10-856-499-1930 Sequence 1930, App
42 98 30.3 176 4 US-10-437-963-164243 Sequence 164243, App
43 97 30.0 181 5 US-10-732-923-450 Sequence 450, App
44 95 29.4 217 4 US-10-425-115-197783 Sequence 197783, App
45 94 29.1 114 4 US-10-425-115-246430 Sequence 246430, App

ALIGNMENTS

RESULT 1
US-09-816-277-2
; Sequence 2, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-09-816-277-2

Query Match 100.0%; Score 323; DB 3; Length 451;
Best Local Similarity 100.0%; Pred. No. 8.9e-24;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EKKPVKRSWGQVLPEPTNLPPKRAKTEDEKEORRVERVLNRRAAQSRRKRLV 60
DB 84 EKKPVKRSWGQVLPEPTNLPPKRAKTEDEKEORRVERVLNRRAAQSRRKRLV 143
QY 61 EALE 64
DB 144 EALE 147

RESULT 2
US-09-816-277-5
; Sequence 5, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells

; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 451
; TYPE: PR1
; ORGANISM: Trichoderma reesei
US-09-816-277-5

Query Match 100.0%; Score 323; DB 3; Length 451;
Best Local Similarity 100.0%; Pred. No. 8.9e-24;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKKPVKKRKSQGVLPEPKTNLP1PRKRAKTEDEKEQRRVERVLNRNRAAQSRRKRLEV 60
DB 84 EKKPVKKRKSQGVLPEPKTNLP1PRKRAKTEDEKEQRRVERVLNRNRAAQSRRKRLEV 143
QY 61 EALE 64
DB 144 EALE 147

RESULT 3

US-10-663-450-2

; Sequence 2, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 451
; TYPE: PR1
; ORGANISM: Trichoderma reesei
US-10-663-450-2

Query Match 100.0%; Score 323; DB 4; Length 451;
Best Local Similarity 100.0%; Pred. No. 8.9e-24;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKKPVKKRKSQGVLPEPKTNLP1PRKRAKTEDEKEQRRVERVLNRNRAAQSRRKRLEV 60
DB 84 EKKPVKKRKSQGVLPEPKTNLP1PRKRAKTEDEKEQRRVERVLNRNRAAQSRRKRLEV 143
QY 61 EALE 64
DB 144 EALE 147

RESULT 4

US-10-663-450-5

; Sequence 5, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.

; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 451
; TYPE: PR1
; ORGANISM: Trichoderma reesei
US-10-663-450-5

Query Match 100.0%; Score 323; DB 4; Length 451;
Best Local Similarity 100.0%; Pred. No. 8.9e-24;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKKPVKKRKSQGVLPEPKTNLP1PRKRAKTEDEKEQRRVERVLNRNRAAQSRRKRLEV 60
DB 84 EKKPVKKRKSQGVLPEPKTNLP1PRKRAKTEDEKEQRRVERVLNRNRAAQSRRKRLEV 143
QY 61 EALE 64
DB 144 EALE 147

RESULT 5

US-09-533-029-104

; Sequence 104, Application US/09533029
; Publication No. US20030046723A1
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Broun, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Keddle, James
; APPLICANT: Pineda, Omaira
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Reuber, Lynne
; TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES
; FILE REFERENCE: MBI-010
; CURRENT APPLICATION NUMBER: US/09/533,029
; CURRENT FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 60/125,814
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 104
; LENGTH: 409
; TYPE: PR1
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1034
US-09-533-029-104

Query Match 95.0%; Score 307; DB 3; Length 409;
Best Local Similarity 93.8%; Pred. No. 3.1e-22;
Matches 60; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKKPVKKRKSQGVLPEPKTNLP1PRKRAKTEDEKEQRRVERVLNRNRAAQSRRKRLEV 60
DB 66 EKKPVKKRKSQGVLPEPKTNLP1PRKRAKTEDEKEQRRVERVLNRNRAAQSRRKRLEV 125
QY 61 EALE 64
DB 144 EALE 147

Db 126 EALE 129

RESULT 6

US-10-295-403-158

Sequence 158, Application US/10295403

Publication No. US20030101481A1

GENERAL INFORMATION:

APPLICANT: Heard, Jacqueline

APPLICANT: Riechmann, Jose Luis

APPLICANT: Adam, Luc

APPLICANT: Broun, Pierre

APPLICANT: Pineda, Omaira

APPLICANT: Reuber, Lynne

APPLICANT: Jiang, Cai-Zhong

APPLICANT: Keddle, James

APPLICANT: Zhang, James

APPLICANT: Benito, Maria-Ines

APPLICANT: Yu, Guo-Liang

APPLICANT: Fromm, Mike

TITLE OF INVENTION: PLANT GENE SEQUENCES I

FILE REFERENCE: MBI-0003

CURRENT APPLICATION NUMBER: US/10/295,403

CURRENT FILING DATE: 2002-11-15

PRIOR APPLICATION NUMBER: US/09/394,519

PRIOR FILING DATE: 1999-09-13

PRIOR APPLICATION NUMBER: 60/101,349

PRIOR FILING DATE: 1998-09-22

PRIOR APPLICATION NUMBER: 60/103,312

PRIOR FILING DATE: 1998-10-06

PRIOR APPLICATION NUMBER: 60/108,734

PRIOR FILING DATE: 1998-11-17

PRIOR APPLICATION NUMBER: 60/113,409

PRIOR FILING DATE: 1998-12-22

NUMBER OF SEQ ID NOS: 170

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 158

LENGTH: 409

TYPE: PRT

ORGANISM: Arabidopsis thaliana

FEATURE:

OTHER INFORMATION: G1034

US-10-295-403-158

Query Match 95.0%; Score 307; DB 4; Length 409;

Best Local Similarity 93.8%; Pred. No. 3.1e-22;

Matches 60; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EKPPVKRKSQGVLPPEKTNLPKRKAKTEDEKQRRVRLNRRRAAQSRRKRQEV 60

Db 66 EKPIKRRKSWGQOLPEKTNLPKRKAKTQDEKQRRVRLNRRRAAQSRRKRQEV 125

Qy 61 EALE 64

Db 126 EALE 129

RESULT 7

US-10-412-699B-684

Sequence 684, Application US/10412699B

Publication No. US20040045049A1

GENERAL INFORMATION:

APPLICANT: Mendel Biotechnology, Inc.

APPLICANT: Zhang, James

APPLICANT: Fromm, Michael E.

APPLICANT: Heard, Jacqueline E.

APPLICANT: Riechmann, Jose Luis

APPLICANT: Adam, Luc J.

APPLICANT: Broun, Pierre E.

APPLICANT: Pineda, Omaira

APPLICANT: Reuber, T. Lynne

APPLICANT: Keddle, James S.

APPLICANT: Yu, Guo-Liang

TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells

FILE REFERENCE: GC590-2

CURRENT APPLICATION NUMBER: US/09/816,277

CURRENT FILING DATE: 2001-03-23

PRIOR APPLICATION NUMBER: US 09/534,692

PRIOR FILING DATE: 2000-03-24

Query Match 95.0%; Score 307; DB 4; Length 409;

Best Local Similarity 93.8%; Pred. No. 3.1e-22;

Matches 60; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EKPPVKRKSQGVLPPEKTNLPKRKAKTEDEKQRRVRLNRRRAAQSRRKRQEV 60

Db 66 EKPIKRRKSWGQOLPEKTNLPKRKAKTQDEKQRRVRLNRRRAAQSRRKRQEV 125

Qy 61 EALE 64

Db 126 EALE 129

RESULT 8

US-09-816-277-19

Sequence 19, Application US/09816277

Publication No. US20010034045A1

GENERAL INFORMATION:

APPLICANT: Penttila, Merja E.

APPLICANT: Ward, Michael

APPLICANT: Wang, Huaming

APPLICANT: Valkonen, Mari J.

APPLICANT: Saloheimo, Markku

TITLE OF INVENTION: Increased Production of Secreted

FILE REFERENCE: MBI-0048CIP

CURRENT APPLICATION NUMBER: US/10/412,699B

CURRENT FILING DATE: 2003-04-10

PRIOR APPLICATION NUMBER: 09/394,519

PRIOR FILING DATE: 1999-09-13

PRIOR APPLICATION NUMBER: 09/489,376

PRIOR FILING DATE: 2000-01-21

PRIOR APPLICATION NUMBER: 09/506,720

PRIOR FILING DATE: 2000-02-17

PRIOR APPLICATION NUMBER: 09/533,030

PRIOR FILING DATE: 2000-03-22

PRIOR APPLICATION NUMBER: 09/533,392

PRIOR FILING DATE: 2000-03-22

PRIOR APPLICATION NUMBER: 09/533,029

PRIOR FILING DATE: 2000-03-22

PRIOR APPLICATION NUMBER: 09/532,591

PRIOR FILING DATE: 2000-03-22

PRIOR APPLICATION NUMBER: 09/533,648

PRIOR FILING DATE: 2000-03-22

PRIOR APPLICATION NUMBER: 09/713,994

PRIOR FILING DATE: 2000-11-16

PRIOR APPLICATION NUMBER: 09/819,142

PRIOR FILING DATE: 2001-03-27

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 2011

SOFTWARE: PatentIn version 3.2

SEQ ID NO 684

LENGTH: 409

TYPE: PRT

ORGANISM: Arabidopsis thaliana

FEATURE:

OTHER INFORMATION: G1034

US-10-412-699B-684

Query Match 95.0%; Score 307; DB 4; Length 409;

Best Local Similarity 93.8%; Pred. No. 3.1e-22;

Matches 60; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EKPPVKRKSQGVLPPEKTNLPKRKAKTEDEKQRRVRLNRRRAAQSRRKRQEV 60

Db 66 EKPIKRRKSWGQOLPEKTNLPKRKAKTQDEKQRRVRLNRRRAAQSRRKRQEV 125

Qy 61 EALE 64

Db 126 EALE 129

```
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-09-816-277-19

Query Match
Best Local Similarity 90.4%; Score 292; DB 3; Length 342;
Matches 57; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 EKKPVKKRKSQGVLPPEPKTNLPPrKRKTEDEKEQRRVRLRNRAAQSRRKRLEV 60
Db 46 EKKPVKKRKSQGVLPPEPKTNLPPrKRKTEDEKEQRRVRLRNRAAQSRRKRLEM 105

QY 61 EALE 64
Db 106 EKLE 109

RESULT 9
US-10-663-450-19
; Sequence 19, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; PRIOR FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-10-663-450-19

Query Match
Best Local Similarity 90.4%; Score 292; DB 4; Length 342;
Matches 57; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 EKKPVKKRKSQGVLPPEPKTNLPPrKRKTEDEKEQRRVRLRNRAAQSRRKRLEV 60
Db 46 EKKPVKKRKSQGVLPPEPKTNLPPrKRKTEDEKEQRRVRLRNRAAQSRRKRLEM 105

QY 61 EALE 64
Db 106 EKLE 109

RESULT 10
US-09-816-277-16
; Sequence 16, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
```

```
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-09-816-277-16

Query Match
Best Local Similarity 90.4%; Score 292; DB 3; Length 386;
Matches 57; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 EKKPVKKRKSQGVLPPEPKTNLPPrKRKTEDEKEQRRVRLRNRAAQSRRKRLEV 60
Db 90 EKKPVKKRKSQGVLPPEPKTNLPPrKRKTEDEKEQRRVRLRNRAAQSRRKRLEM 149

QY 61 EALE 64
Db 150 EKLE 153

RESULT 11
US-10-663-450-16
; Sequence 16, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; PRIOR FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-10-663-450-16

Query Match
Best Local Similarity 90.4%; Score 292; DB 4; Length 386;
Matches 57; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 EKKPVKKRKSQGVLPPEPKTNLPPrKRKTEDEKEQRRVRLRNRAAQSRRKRLEV 60
Db 90 EKKPVKKRKSQGVLPPEPKTNLPPrKRKTEDEKEQRRVRLRNRAAQSRRKRLEM 149

QY 61 EALE 64
Db 150 EKLE 153

RESULT 12
US-09-816-277-4
; Sequence 4, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
```


; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-816-277-4

Query Match 89.2%; Score 288; DB 3; Length 349;
Best Local Similarity 87.5%; Pred. No. 2e-20;
Matches 56; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 EKKPVKKRSGQVLPEPKTNLPKRKRAKTEDEKQRRVRLNRRAAQSRRKRLEV 60
DB 52 EKKPAKRRSGQELPVPKTNLPKRKRAKTEDEKQRRVRLNRRAAQTSSRRKRLEM 111
QY 61 EALE 64
DB 112 EKLE 115

RESULT 13
US-09-816-277-6
; Sequence 6, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-816-277-6

Query Match 89.2%; Score 288; DB 3; Length 349;
Best Local Similarity 87.5%; Pred. No. 2e-20;
Matches 56; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 EKKPVKKRSGQVLPEPKTNLPKRKRAKTEDEKQRRVRLNRRAAQSRRKRLEV 60
DB 52 EKKPAKRRSGQELPVPKTNLPKRKRAKTEDEKQRRVRLNRRAAQTSSRRKRLEM 111
QY 61 EALE 64
DB 112 EKLE 115

RESULT 14
US-10-663-450-4
; Sequence 4, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming

; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-663-450-4

Query Match 89.2%; Score 288; DB 4; Length 349;
Best Local Similarity 87.5%; Pred. No. 2e-20;
Matches 56; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 EKKPVKKRSGQVLPEPKTNLPKRKRAKTEDEKQRRVRLNRRAAQSRRKRLEV 60
DB 52 EKKPAKRRSGQELPVPKTNLPKRKRAKTEDEKQRRVRLNRRAAQTSSRRKRLEM 111
QY 61 EALE 64
DB 112 EKLE 115

RESULT 15
US-10-663-450-6
; Sequence 6, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-663-450-6

Query Match 89.2%; Score 288; DB 4; Length 349;
Best Local Similarity 87.5%; Pred. No. 2e-20;
Matches 56; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 EKKPVKKRSGQVLPEPKTNLPKRKRAKTEDEKQRRVRLNRRAAQSRRKRLEV 60
DB 52 EKKPAKRRSGQELPVPKTNLPKRKRAKTEDEKQRRVRLNRRAAQTSSRRKRLEM 111
QY 61 EALE 64
DB 112 EKLE 115

Search completed: November 22, 2005, 15:54:35
Job time : 117 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2005, 15:33:00 ; Search time 147.5 Seconds
(without alignments)
306.128 Million cell updates/sec

Title: US-10-663-450-5_COPY_84_147
Perfect score: 323
Sequence: 1 EKKPVKRSWGQVLPEPKT.....NRRAAQSRKRKRLEVALE 64

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	323	100.0	451	2	Q8TFP3_TRIRE	Q8tfp3 trichoderma
2	319	98.8	430	2	Q7SHF0_NEUCR	Q7shf0 neurospora
3	296	91.6	552	2	Q51KW8_MAGGR	Q51kw8 magnaporthe
4	292	90.4	342	2	Q6W8X2_ASPNG	Q6w8x2 aspergillus
5	288	89.2	347	2	Q5AQN3_EMENI	Q5aqn3 aspergillus
6	288	89.2	350	2	Q8TFU8_EMENI	Q8tfu8 emericella
7	286	88.5	429	2	Q4HTT5_GIBZE	Q4htt5 gibberella
8	285	88.2	433	2	Q4WEY8_ASPFU	Q4wey8 aspergillus
9	178	55.1	260	2	Q6BQC2_DEBHA	Q6bqc2 debaryomyce
10	175	54.3	273	2	Q6CKQ1_KLUJA	Q6ckq1 kluyveromyce
11	175	54.3	299	2	Q6CEV1_YARLI	Q6cev1 yarrowia li
12	175	54.2	230	1	HAC1_YEAS	P41546 saccharomyc
13	174	53.9	357	2	Q5AA52_CANAL	Q5aa52 candida alb
14	168	52.0	228	2	Q75BQ5_ASHGO	Q75bq5 ashbya goss
15	148	45.8	329	2	Q6FLY3_CANGA	Q6fly3 candida gla
16	107	33.1	336	2	Q5VR11_ORYSA	Q5vr11 oryza sativ
17	106	32.8	168	1	HY5_ARATH	O24646 arabidopsis
18	102	31.6	69	2	Q58L60_BRACM	Q58l60 brassica ca
19	102	31.6	109	2	Q84XX8_ERARP	Q84xx8 brassica ra
20	101	31.3	158	1	HY5_LYCES	Q9sm50 lycopersico
21	100.5	31.1	141	2	Q677A7_9ASPA	Q677a7 hyacinthus
22	98	30.3	176	2	Q69XK6_ORYSA	Q69xk6 oryza sativ
23	96	29.7	252	2	Q6NX18_XENTR	Q6nxt8 xenopus tro
24	94	29.1	188	2	Q6ZHT8_ORYSA	Q6zht8 oryza sativ
25	94	29.1	321	2	Q6GRY7_LOTJA	Q6gry7 lotus japon
26	94	29.1	646	2	Q6AU90_ORYSA	Q6au90 oryza sativ
27	94	29.1	686	2	Q61D33_CAENR	Q61d33 caenorhabdi
28	93	28.8	120	2	Q682B6_ARATH	Q682b6 arabidopsis
29	93	28.8	149	1	HYH_ARATH	Q8w191 arabidopsis
30	92	28.5	350	2	Q90ZR7_XENLA	Q90zr7 xenopus lae
31	92	28.5	396	2	Q7ZYC2_XENLA	Q7zyc2 xenopus lae

RESULT 1									
Q8TFP3	Q8TFP3	Q8TFP3	Q8TFP3	Q8TFP3	Q8TFP3	Q8TFP3	Q8TFP3	Q8TFP3	Q8TFP3
ID	ID	ID	ID	ID	ID	ID	ID	ID	ID
Q8TFP3	Q8TFP3	Q8TFP3	Q8TFP3	Q8TFP3	Q8TFP3	Q8TFP3	Q8TFP3	Q8TFP3	Q8TFP3
AC	AC	AC	AC	AC	AC	AC	AC	AC	AC
DT	DT	DT	DT	DT	DT	DT	DT	DT	DT
01-JUN-2002	01-JUN-2002	01-JUN-2002	01-JUN-2002	01-JUN-2002	01-JUN-2002	01-JUN-2002	01-JUN-2002	01-JUN-2002	01-JUN-2002
(Tremblrel. 21, Created)	(Tremblrel. 21, Created)	(Tremblrel. 21, Created)	(Tremblrel. 21, Created)	(Tremblrel. 21, Created)	(Tremblrel. 21, Created)	(Tremblrel. 21, Created)	(Tremblrel. 21, Created)	(Tremblrel. 21, Created)	(Tremblrel. 21, Created)
01-JUN-2002	01-JUN-2002	01-JUN-2002	01-JUN-2002	01-JUN-2002	01-JUN-2002	01-JUN-2002	01-JUN-2002	01-JUN-2002	01-JUN-2002
(Tremblrel. 21, Last sequence update)	(Tremblrel. 21, Last sequence update)	(Tremblrel. 21, Last sequence update)	(Tremblrel. 21, Last sequence update)	(Tremblrel. 21, Last sequence update)	(Tremblrel. 21, Last sequence update)	(Tremblrel. 21, Last sequence update)	(Tremblrel. 21, Last sequence update)	(Tremblrel. 21, Last sequence update)	(Tremblrel. 21, Last sequence update)
01-MAR-2004	01-MAR-2004	01-MAR-2004	01-MAR-2004	01-MAR-2004	01-MAR-2004	01-MAR-2004	01-MAR-2004	01-MAR-2004	01-MAR-2004
(Tremblrel. 26, Last annotation update)	(Tremblrel. 26, Last annotation update)	(Tremblrel. 26, Last annotation update)	(Tremblrel. 26, Last annotation update)	(Tremblrel. 26, Last annotation update)	(Tremblrel. 26, Last annotation update)	(Tremblrel. 26, Last annotation update)	(Tremblrel. 26, Last annotation update)	(Tremblrel. 26, Last annotation update)	(Tremblrel. 26, Last annotation update)
Transcription factor.	Transcription factor.	Transcription factor.	Transcription factor.	Transcription factor.	Transcription factor.	Transcription factor.	Transcription factor.	Transcription factor.	Transcription factor.
Name=hacl;	Name=hacl;	Name=hacl;	Name=hacl;	Name=hacl;	Name=hacl;	Name=hacl;	Name=hacl;	Name=hacl;	Name=hacl;
Trichoderma reesei (Hypocrea jecorina).	Trichoderma reesei (Hypocrea jecorina).	Trichoderma reesei (Hypocrea jecorina).	Trichoderma reesei (Hypocrea jecorina).	Trichoderma reesei (Hypocrea jecorina).	Trichoderma reesei (Hypocrea jecorina).	Trichoderma reesei (Hypocrea jecorina).	Trichoderma reesei (Hypocrea jecorina).	Trichoderma reesei (Hypocrea jecorina).	Trichoderma reesei (Hypocrea jecorina).
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.	Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.	Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.	Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.	Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.	Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.	Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.	Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.	Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.	Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
NCBI_TaxID=51453;	NCBI_TaxID=51453;	NCBI_TaxID=51453;	NCBI_TaxID=51453;	NCBI_TaxID=51453;	NCBI_TaxID=51453;	NCBI_TaxID=51453;	NCBI_TaxID=51453;	NCBI_TaxID=51453;	NCBI_TaxID=51453;
[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]
NUCLEOTIDE SEQUENCE.	NUCLEOTIDE SEQUENCE.	NUCLEOTIDE SEQUENCE.	NUCLEOTIDE SEQUENCE.	NUCLEOTIDE SEQUENCE.	NUCLEOTIDE SEQUENCE.	NUCLEOTIDE SEQUENCE.	NUCLEOTIDE SEQUENCE.	NUCLEOTIDE SEQUENCE.	NUCLEOTIDE SEQUENCE.
Salohelmo M.L.A., Valkonen M., Penttilae M.E.;	Salohelmo M.L.A., Valkonen M., Penttilae M.E.;	Salohelmo M.L.A., Valkonen M., Penttilae M.E.;	Salohelmo M.L.A., Valkonen M., Penttilae M.E.;	Salohelmo M.L.A., Valkonen M., Penttilae M.E.;	Salohelmo M.L.A., Valkonen M., Penttilae M.E.;	Salohelmo M.L.A., Valkonen M., Penttilae M.E.;	Salohelmo M.L.A., Valkonen M., Penttilae M.E.;	Salohelmo M.L.A., Valkonen M., Penttilae M.E.;	Salohelmo M.L.A., Valkonen M., Penttilae M.E.;
Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.	Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.	Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.	Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.	Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.	Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.	Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.	Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.	Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.	Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
EMBL; A413272; CAC88374.1; --; Genomic_DNA.	EMBL; A413272; CAC88374.1; --; Genomic_DNA.	EMBL; A413272; CAC88374.1; --; Genomic_DNA.	EMBL; A413272; CAC88374.1; --; Genomic_DNA.	EMBL; A413272; CAC88374.1; --; Genomic_DNA.	EMBL; A413272; CAC88374.1; --; Genomic_DNA.	EMBL; A413272; CAC88374.1; --; Genomic_DNA.	EMBL; A413272; CAC88374.1; --; Genomic_DNA.	EMBL; A413272; CAC88374.1; --; Genomic_DNA.	EMBL; A413272; CAC88374.1; --; Genomic_DNA.
GO; GO:0005634; C:nucleus; IEA.	GO; GO:0005634; C:nucleus; IEA.	GO; GO:0005634; C:nucleus; IEA.	GO; GO:0005634; C:nucleus; IEA.	GO; GO:0005634; C:nucleus; IEA.	GO; GO:0005634; C:nucleus; IEA.	GO; GO:0005634; C:nucleus; IEA.	GO; GO:0005634; C:nucleus; IEA.	GO; GO:0005634; C:nucleus; IEA.	GO; GO:0005634; C:nucleus; IEA.
GO; GO:0003677; F:DNA binding; IEA.	GO; GO:0003677; F:DNA binding; IEA.	GO; GO:0003677; F:DNA binding; IEA.	GO; GO:0003677; F:DNA binding; IEA.	GO; GO:0003677; F:DNA binding; IEA.	GO; GO:0003677; F:DNA binding; IEA.	GO; GO:0003677; F:DNA binding; IEA.	GO; GO:0003677; F:DNA binding; IEA.	GO; GO:0003677; F:DNA binding; IEA.	GO; GO:0003677; F:DNA binding; IEA.
GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
InterPro; IPR011700; bZIP 2.	InterPro; IPR011700; bZIP 2.	InterPro; IPR011700; bZIP 2.	InterPro; IPR011700; bZIP 2.	InterPro; IPR011700; bZIP 2.	InterPro; IPR011700; bZIP 2.	InterPro; IPR011700; bZIP 2.	InterPro; IPR011700; bZIP 2.	InterPro; IPR011700; bZIP 2.	InterPro; IPR011700; bZIP 2.
InterPro; IPR004827; TF_bZIP.	InterPro; IPR004827; TF_bZIP.	InterPro; IPR004827; TF_bZIP.	InterPro; IPR004827; TF_bZIP.	InterPro; IPR004827; TF_bZIP.	InterPro; IPR004827; TF_bZIP.	InterPro; IPR004827; TF_bZIP.	InterPro; IPR004827; TF_bZIP.	InterPro; IPR004827; TF_bZIP.	InterPro; IPR004827; TF_bZIP.
Pfam; PF07716; bZIP_2; 1.	Pfam; PF07716; bZIP_2; 1.	Pfam; PF07716; bZIP_2; 1.	Pfam; PF07716; bZIP_2; 1.	Pfam; PF07716; bZIP_2; 1.	Pfam; PF07716; bZIP_2; 1.	Pfam; PF07716; bZIP_2; 1.	Pfam; PF07716; bZIP_2; 1.	Pfam; PF07716; bZIP_2; 1.	Pfam; PF07716; bZIP_2; 1.
SMART; SM00338; BRLZ; 1.	SMART; SM00338; BRLZ; 1.	SMART; SM00338; BRLZ; 1.	SMART; SM00338; BRLZ; 1.	SMART; SM00338; BRLZ; 1.	SMART; SM00338; BRLZ; 1.	SMART; SM00338; BRLZ; 1.	SMART; SM00338; BRLZ; 1.	SMART; SM00338; BRLZ; 1.	SMART; SM00338; BRLZ; 1.
PROSITE; PS0217; BZIP; 1.	PROSITE; PS0217; BZIP; 1.	PROSITE; PS0217; BZIP; 1.	PROSITE; PS0217; BZIP; 1.	PROSITE; PS0217; BZIP; 1.	PROSITE; PS0217; BZIP; 1.	PROSITE; PS0217; BZIP; 1.	PROSITE; PS0217; BZIP; 1.	PROSITE; PS0217; BZIP; 1.	PROSITE; PS0217; BZIP; 1.
DNA-binding; Nuclear protein.	DNA-binding; Nuclear protein.	DNA-binding; Nuclear protein.	DNA-binding; Nuclear protein.	DNA-binding; Nuclear protein.	DNA-binding; Nuclear protein.	DNA-binding; Nuclear protein.	DNA-binding; Nuclear protein.	DNA-binding; Nuclear protein.	DNA-binding; Nuclear protein.
SEQUENCE 451 AA; 49277 MW; 600F10B471EA3AD3 CRC64;	SEQUENCE 451 AA; 49277 MW; 600F10B471EA3AD3 CRC64;	SEQUENCE 451 AA; 49277 MW; 600F10B471EA3AD3 CRC64;	SEQUENCE 451 AA; 49277 MW; 600F10B471EA3AD3 CRC64;	SEQUENCE 451 AA; 49277 MW; 600F10B471EA3AD3 CRC64;	SEQUENCE 451 AA; 49277 MW; 600F10B471EA3AD3 CRC64;	SEQUENCE 451 AA; 49277 MW; 600F10B471EA3AD3 CRC64;	SEQUENCE 451 AA; 49277 MW; 600F10B471EA3AD3 CRC64;	SEQUENCE 451 AA; 49277 MW; 600F10B471EA3AD3 CRC64;	SEQUENCE 451 AA; 49277 MW; 600F10B471EA3AD3 CRC64;
Query Match 100.0%; Score 323; DB 2; Length 451;	Query Match 100.0%; Score 323; DB 2; Length 451;	Query Match 100.0%; Score 323; DB 2; Length 451;	Query Match 100.0%; Score 323; DB 2; Length 451;	Query Match 100.0%; Score 323; DB 2; Length 451;	Query Match 100.0%; Score 323; DB 2; Length 451;	Query Match 100.0%; Score 323; DB 2; Length 451;	Query Match 100.0%; Score 323; DB 2; Length 451;	Query Match 100.0%; Score 323; DB 2; Length 451;	Query Match 100.0%; Score 323; DB 2; Length 451;
Best Local Similarity 100.0%; Pred. No. 5.1e-21;	Best Local Similarity 100.0%; Pred. No. 5.1e-21;	Best Local Similarity 100.0%; Pred. No. 5.1e-21;	Best Local Similarity 100.0%; Pred. No. 5.1e-21;	Best Local Similarity 100.0%; Pred. No. 5.1e-21;	Best Local Similarity 100.0%; Pred. No. 5.1e-21;	Best Local Similarity 100.0%; Pred. No. 5.1e-21;	Best Local Similarity 100.0%; Pred. No. 5.1e-21;	Best Local Similarity 100.0%; Pred. No. 5.1e-21;	Best Local Similarity 100.0%; Pred. No. 5.1e-21;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EKKPVKRSWGQVLPEPKTLPKRKRAKTEKEQRRVERVLNRRAAQSRRKRLEV 60	QY 1 EKKPVKRSWGQVLPEPKTLPKRKRAKTEKEQRRVERVLNRRAAQSRRKRLEV 60	QY 1 EKKPVKRSWGQVLPEPKTLPKRKRAKTEKEQRRVERVLNRRAAQSRRKRLEV 60	QY 1 EKKPVKRSWGQVLPEPKTLPKRKRAKTEKEQRRVERVLNRRAAQSRRKRLEV 60	QY 1 EKKPVKRSWGQVLPEPKTLPKRKRAKTEKEQRRVERVLNRRAAQSRRKRLEV 60	QY 1 EKKPVKRSWGQVLPEPKTLPKRKRAKTEKEQRRVERVLNRRAAQSRRKRLEV 60	QY 1 EKKPVKRSWGQVLPEPKTLPKRKRAKTEKEQRRVERVLNRRAAQSRRKRLEV 60	QY 1 EKKPVKRSWGQVLPEPKTLPKRKRAKTEKEQRRVERVLNRRAAQSRRKRLEV 60	QY 1 EKKPVKRSWGQVLPEPKTLPKRKRAKTEKEQRRVERVLNRRAAQSRRKRLEV 60	QY 1 EKKPVKRSWGQVLPEPKTLPKRKRAKTEKEQRRVERVLNRRAAQSRRKRLEV 60
DB 84 EKKPVKRSWGQVLPEPKTLPKRKRAKTEKEQRRVERVLNRRAAQSRRKRLEV 143	DB 84 EKKPVKRSWGQVLPEPKTLPKRKRAKTEKEQRRVERVLNRRAAQSRRKRLEV 143	DB 84 EKKPVKRSWGQVLPEPKTLPKRKRAKTEKEQRRVERVLNRRAAQSRRKRLEV 143	DB 84 EKKPVKRSWGQVLPEPKTLPKRKRAKTEKEQRRVERVLNRRAAQSRRKRLEV 143	DB 84 EKKPVKRSWGQVLPEPKTLPKRKRAKTEKEQRRVERVLNRRAAQSRRKRLEV 143	DB 84 EKKPVKRSWGQVLPEPKTLPKRKRAKTEKEQRRVERVLNRRAAQSRRKRLEV 143	DB 84 EKKPVKRSWGQVLPEPKTLPKRKRAKTEKEQRRVERVLNRRAAQSRRKRLEV 143	DB 84 EKKPVKRSWGQVLPEPKTLPKRKRAKTEKEQRRVERVLNRRAAQSRRKRLEV 143	DB 84 EKKPVKRSWGQVLPEPKTLPKRKRAKTEKEQRRVERVLNRRAAQSRRKRLEV 143	DB 84 EKKPVKRSWGQVLPEPKTLPKRKRAKTEKEQRRVERVLNRRAAQSRRKRLEV 143
QY 61 EALE 64	QY 61 EALE 64	QY 61 EALE 64	QY 61 EALE 64	QY 61 EALE 64	QY 61 EALE 64	QY 61 EALE 64	QY 61 EALE 64	QY 61 EALE 64	QY 61 EALE 64
DB 144 EALE 147	DB 144 EALE 147	DB 144 EALE 147	DB 144 EALE 147	DB 144 EALE 147	DB 144 EALE 147	DB 144 EALE 147	DB 144 EALE 147	DB 144 EALE 147	DB 144 EALE 147
RESULT 2									
Q7SHF0	Q7SHF0	Q7SHF0	Q7SHF0	Q7SHF0	Q7SHF0	Q7SHF0	Q7SHF0	Q7SHF0	Q7SHF0
ID	ID	ID	ID	ID	ID	ID	ID	ID	ID
Q7SHF0	Q7SHF0	Q7SHF0	Q7SHF0	Q7SHF0	Q7SHF0	Q7SHF0	Q7SHF0	Q7SHF0	Q7SHF0
AC	AC	AC	AC	AC	AC	AC	AC	AC	AC
Q7SHF0	Q7SHF0	Q7SHF0	Q7SHF0	Q7SHF0	Q7SHF0	Q7SHF0	Q7SHF0	Q7SHF0	Q7SHF0
DT	DT	DT	DT	DT	DT	DT	DT	DT	DT
01-MAR-2004	01-MAR-2004	01-MAR-2004	01-MAR-2004	01-MAR-2004	01-MAR-2004	01-MAR-2004	01-MAR-2004	01-MAR-2004	01-MAR-2004
(Tremblrel. 26, Created)	(Tremblrel. 26, Created)	(Tremblrel. 26, Created)	(Tremblrel. 26, Created)	(Tremblrel. 26, Created)	(Tremblrel. 26, Created)	(Tremblrel. 26, Created)	(Tremblrel. 26, Created)	(Tremblrel. 26, Created)	(Tremblrel. 26, Created)
01-MAR-2004	01-MAR-2004	01-MAR-2004	01-MAR-2004	01-MAR-2004	01-MAR-2004	01-MAR-2004	01-MAR-2004	01-MAR-2004	01-MAR-2004
(Tremblrel. 26, Last sequence update)	(Tremblrel. 26, Last sequence update)	(Tremblrel. 26, Last sequence update)	(Tremblrel. 26, Last sequence update)	(Tremblrel. 26, Last sequence update)	(Tremblrel. 26, Last sequence update)	(Tremblrel. 26, Last sequence update)	(Tremblrel. 26, Last sequence update)	(Tremblrel. 26, Last sequence update)	(Tremblrel. 26, Last sequence update)
Predicted protein.	Predicted protein.	Predicted protein.	Predicted protein.	Predicted protein.	Predicted protein.	Predicted protein.	Predicted protein.	Predicted protein.	Predicted protein.
Name=NCU01856.1;	Name=NCU01856.1;	Name=NCU01856.1;	Name=NCU01856.1;	Name=NCU01856.1;	Name=NCU01856.1;	Name=NCU01856.1;	Name=NCU01856.1;	Name=NCU01856.1;	Name=NCU01856.1;
Neurospora crassa.	Neurospora crassa.	Neurospora crassa.	Neurospora crassa.	Neurospora crassa.	Neurospora crassa.	Neurospora crassa.	Neurospora crassa.	Neurospora crassa.	Neurospora crassa.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.	Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.	Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.	Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.	Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.	Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.	Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.	Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.	Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.	Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.

Q93xa0 phaseolus v
Q39896 glycine max
Q39895 glycine max
O44743 caenorhabdi
Q61817 mus musculus
Q5tcv1 homo sapien
Q5r5z9 pongo pygma
O43889 homo sapien
Q5dtk2 schistosoma
Q569tc3 xenopus lae
O04234 vicia faba
Q98rx1 guillardia
Q4p8q4 ustilago ma
Q54y73 dictyosteli

```

OX NCBI_TaxID=5141;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., Fitzhugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Selltreinikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Newes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamysseelis M., Mauceli E., Bielke C., Rudd S., Frishman D.,
RA Kryzstofova S., Rasmussen C., Metzenberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Osmari S.A.,
RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbold D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.;
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL Nature 0:0-0(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AABX0100004; EAA36251.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; bZIP_2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07716; bZIP_2; 1.
DR PROSITE; PS50217; bZIP; 1.
SQ SEQUENCE 430 AA; 46599 MW; 0EE0657CFA6160DA CRC64;

Query Match 98.8%; Score 319; DB 2; Length 430;
Best Local Similarity 98.4%; Pred. No. 1.1e-20;
Matches 63; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EKKPVKKRSGQVLPKPTNLPKRKATDEKEQRRVERVLNRRAAQSRRKRLEVE 60
DB 87 EKKPVKKRSGQVLPKPTNLPKRKATDEKEQRRVERVLNRRAAQSRRKRLEVE 146

OY 61 EALE 64
DB 147 EGLE 150

RESULT 3
OY Q51KW8_MAGGR PRELIMINARY; PRT; 556 AA.
AC O51KW8;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
OS ORFNames=MG09010.4;
GN Magnaporthe grisea 70-15.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
OX NCBI_TaxID=242507;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Birren B., Nussbaum C., Abebe A., Abouelleil A., Adekova E.,
RA Ait-zahra M., Allen N., Allen T., An P., Anderson M., Anderson S.,
RA Arachchi H., Armbruster J., Bachantsang P., Baldwin J., Barry A.,
RA Bayul T., Blitshsteyn B., Bloom T., Blye J., Boguslavskiy L.,
RA Borowsky M., Boukgalter B., Brunache A., Butler J., Calixte N.,
RA Calvo S., Camarata J., Campo K., Chang J., Cheshatsang Y., Citroen M.,
RA Collimore A., Considine T., Cook A., Cooke P., Corum B., Cuomo C.,
RA David R., Dawoe T., Degray S., Dodge S., Dooley K., Dorje P.,
RA Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T., Engels R.,
RA Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,
RA Fitzgerald M., Foley K., Gage D., Galagan J., Gearin G., Gnerre S.,

```

```

RA Gnrirke A., Goyette A., Graham J., Grandbois E., Gyaltsen K., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,
RA Honan T., Horn A., Houde N., Hughes L., Hulme W., Huseby E., Iliev I.,
RA Jaffe D., Jones C., Kamal M., Kamat A., Kamysseelis M., Karlsson E.,
RA Kells C., Kieu A., Kishner P., Kodira C., Kulbokas E., Labutti K.,
RA Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,
RA Lindblad-toh K., Liu X., Lokyitsang T., Lokyitsang Y., Lucien O.,
RA Lui A., Ma L.J., Mabbitt R., Macdonald J., Maclean C., Major J.,
RA Manning J., Marabella R., Maru K., Matthews C., Mauceli E.,
RA McCarthy M., McDonough S., Mcghee T., Meldrim J., Meneus L.,
RA Mesirov J., Mihalov A., Mihova T., Mikkelson T., Mieneus V., Moru K.,
RA Mozes J., Mulrain L., Munson G., Naylor J., Neues C., Nguyen C.,
RA Nguyen N., Nguyen T., Nicol R., Nielsen C., Nizzari M., Norbu C.,
RA Norbu N., O'donnell P., Okawo O., O'leary S., Omotosho B.,
RA O'Neill K., Osman S., Parker S., Perrin D., Phunkhang P., Pignani B.,
RA Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond C.,
RA Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,
RA Rutnan M., Schupbach R., Seaman C., Settipalli S., Sharpe T.,
RA Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,
RA Spencer B., Stalker J., Stange-thomann N., Stavropoulos S.,
RA Stetson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,
RA Tenney P., Tesfaye S., Theodore J., Thoultsang Y., Topham K.,
RA Towey S., Tsamia T., Tsomo N., Vallee D., Vassiliev H.,
RA Venkataraman V., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,
RA Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,
RA Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembeck L.,
RA Zimmer A., Zody M., Lander E.;
RT "The genome sequence of Magnaporthe grisea.";
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Dean R., Mitchell T., Brown D., Pan H., Thon M.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Zhu H., Blackmon B.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AACU01001606; EAA47880.1; -; Genomic_DNA.
DR InterPro; IPR011700; bZIP_2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07716; bZIP_2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS50217; bZIP; 1.
KW DNA-binding; Hypothetical protein; Nuclear protein.
SQ SEQUENCE 556 AA; 60798 MW; CFFC4C8D2DC4963C CRC64;

Query Match 91.6%; Score 296; DB 2; Length 556;
Best Local Similarity 93.7%; Pred. No. 1.7e-18;
Matches 59; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 KKPVKRKSQVLPKPTNLPKRKATDEKEQRRVERVLNRRAAQSRRKRLEVE 61
DB 88 EKKTKRKSQVLPKPTNLPKRKATDEKEQRRVERVLNRRAAQSRRKRLEVE 147

OY 62 ALE 64
DB 148 ALE 150

RESULT 4
OY Q6W8X2 ASPNG PRELIMINARY; PRT; 342 AA.
ID Q6W8X2;
AC Q6W8X2;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Transcription factor HACA.
GN Name=hacA;

```

```

OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Mulder H.J.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY303684; AAQ73495.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; bZIP_2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07716; bZIP_2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS02117; BZIP; 1.
DR PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 342 AA; 37148 MW; 623B3941A55C9C05 CRC64;

Query Match 90.4%; Score 292; DB 2; Length 342;
Best Local Similarity 89.1%; Pred. No. 2.4e-18;
Matches 57; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 EKKPVKKRSGWGLPEPKTNLPKRKATDEKEQRRVERVLRNRAAQSRRKRLEV 60
DB 46 EKKPVKKRSGWGLPEPKTNLPKRKATDEKEQRRVERVLRNRAAQSRRKRLEV 105
QY 61 EALE 64
DB 106 EKLE 109

RESULT 5
QSAQN3 EMENI PRELIMINARY; PRT; 347 AA.
AC QSAQN3;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DE Hypothetical protein.
GN ORFNames=AN9397.2;
OS Aspergillus nidulans FGSC A4.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=227321;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavskiy L.,
RA Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
RA Choepel Y., Collymore A., Cook A., Cooke P., Corum B., Dearrellano K.,
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
RA Erickson J., Faro S., Ferreira P., FitzGerald M., Gage D., Galagan J.,
RA Gardyna S., Gnirre S., Graham L., Grand-Pierre N., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,
RA Jaffe D., Johnson R., Jones C., Kanat M., Kanat A., Karatas A.,
RA Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
RA Ma L.-J., Mabbitt R., MacLean C., Macdonald P., Major J., Manning J.,
RA Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,
RA Mihova T., Mlenga V., Murphy T., Naylor J., Nguyen C., Nicol R.,
RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neill D.,
RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,
RA Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,
RA Roman J., Schauer S., Schupbach R., Seaman S., Severy P., Smirnov S.,
RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,
RA Talamas J., Tesfaye S., Theodore J., Topham K., Travers M.,
RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
RA Lander E.;
RT "Genome Sequence of Aspergillus nidulans.";
```

```

RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AACD01000172; EAA66464.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
KW DNA-binding; Hypothetical protein; Nuclear protein.
SQ SEQUENCE 347 AA; 37841 MW; A7CAAFEE761B3E1 CRC64;

Query Match 89.2%; Score 288; DB 2; Length 347;
Best Local Similarity 87.5%; Pred. No. 5.5e-18;
Matches 56; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 EKKPVKKRSGWGLPEPKTNLPKRKATDEKEQRRVERVLRNRAAQSRRKRLEV 60
DB 53 EKKPVKKRSGWGLPEPKTNLPKRKATDEKEQRRVERVLRNRAAQSRRKRLEV 112
QY 61 EALE 64
DB 113 EKLE 116

RESULT 6
Q8TFU8 EMENI PRELIMINARY; PRT; 350 AA.
AC Q8TFU8;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Transcription factor.
GN Name=hacA;
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Saloheimo M.L.A., Valkonen M., Penttilae M.E.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ413273; CAC88375.1; -; Genomic_DNA.
DR HSSP; P05412; 1JNN.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; bZIP_2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07716; bZIP_2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS02117; BZIP; 1.
DR PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 350 AA; 38192 MW; 4F78340096064F44 CRC64;

Query Match 89.2%; Score 288; DB 2; Length 350;
Best Local Similarity 87.5%; Pred. No. 5.6e-18;
Matches 56; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 EKKPVKKRSGWGLPEPKTNLPKRKATDEKEQRRVERVLRNRAAQSRRKRLEV 60
DB 53 EKKPVKKRSGWGLPEPKTNLPKRKATDEKEQRRVERVLRNRAAQSRRKRLEV 112
QY 61 EALE 64
DB 113 EKLE 116

RESULT 7
Q4HTT5 GIBZE PRELIMINARY; PRT; 429 AA.
ID Q4HTT5;
AC Q4HTT5;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
```

```

GN ORFNames=FG11623.1;
OS Gibberella zeae PH-1.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
NCBI_TaxID=229533;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PH-1;
RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavsky L.,
RA Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
RA Choepel Y., Collymore A., Cook A., Cooke P., Corum B., DeArellano K.,
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
RA Erickson J., Faro S., Ferreira P., Fitzgerald M., Gage D., Galagan J.,
RA Gardyna S., Gnerre S., Graham L., Grand-Pierre N., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,
RA Jaffe C., Johnson R., Jones C., Kamal M., Kamat A., Karatas A.,
RA Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
RA Ma L.-J., Mabbitt R., Maclean C., Macdonald P., Major J., Manning J.,
RA Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,
RA Mihova T., Mlenga V., Murphy T., Naylor J., Nguyen C., Nicol R.,
RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neil D.,
RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,
RA Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,
RA Roman J., Schauer S., Schuppback R., Seaman S., Severy P., Smirnov S.,
RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,
RA Talamas J., Testaye S., Theodore J., Topham K., Travers M.,
RA Vassiliou H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
RA Lander E.
RT "Fusarium graminearum genome sequence."
RL Submitted (FEB-2004) to the EMBL/GenBank/DBSJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBSJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AACM01000497; EAA78735.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 429 AA; 46630 MW; 04CCDE40BD392901 CRC64;

Query Match 88.5%; Score 286; DB 2; Length 429;
Best Local Similarity 74.4%; Pred. No. 1e-17; 3; Indels 18; Gaps 1;
Matches 61; Conservative 0; Mismatches 0;

Qy 1 EKPKVKRKSQGVLPPEKPTNLPP-----RKRAKTEDEKEQRRVERV 42
Db 82 EKTKTKRKSQGVLPPEKPTNLPRYGDTLRHVVDESFDTLNRKRAKTEDEKEQRRVERV 141
Qy 43 LRNRRAAQSSRRKRQLEVEALE 64
Db 142 LRNRRAAQSSRRKRQLEVEALE 163

RESULT 8
Q4WEY8_ASPPU
ID Q4WEY8 ASPPU PRELIMINARY; PRT; 433 AA.
AC Q4WEY8;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE BZIP transcription factor (HACA), putative.
GN ORFNames=Afu3g04070;
OS Aspergillus fumigatus Af293.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
NCBI_TaxID=330879;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Af293;
RA Nierman W., Pain A., Anderson M.J., Wortman J., Kim H., Stanley J.,
RA Arroya J., Berriman M., Abe K., Archer D.B., Bermejo C., Bennett J.,
RA Bowyer P., Chen D., Collins M., Coulson R., Davies R., Dyer P.S.,
RA Farman M., Fedorova N., Fedorova N., Feldblyum T.V., Fischer R.,
RA Fosker N., Fraser A., Garcia J.L., Garcia M.J., Goble A.,

```

Best Local Similarity 73.9%; Pred. No. 3.4e-08;
Matches 34; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Qy 19 KTNLPKPKRAKTDEKEQRVERVLNRRAAQSRRKKLEVALE 64
|||
Db 15 KPTLPKPKRAKTQEKEQRRIEILNRRANHQSRKKRLHVORLE 60
|||

RESULT 11
Q6CEV1_YARL1 PRELIMINARY; PRT; 299 AA.
AC Q6CEV1; 28, Created)
DT 23-OCT-2004 (T-EMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (T-EMBLrel. 28, Last annotation update)
DT 25-OCT-2004 (T-EMBLrel. 28, Last annotation update)
De Yarrowia lipolytica chromosome B of strain CLIB99 of Yarrowia
lipolytica.
GN OrderedLocusNames=YALI0B12716g;
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
NCBI_TaxID=4952;
[1]
RN NCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP STRAIN=CLIB 122 / E 150;
RX PubMed=15229592; DOI=10.1038/nature02579;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
R Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
R Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
R Barney S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
R Boixrame A., Boyer J.J., Cattolico L., Confantoulet F., de Daruvar A.,
R Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
R Kerstaeve F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
R Karst A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
R Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulou O.,
R Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,
R Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
R Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
R Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
R Wincker P., Souciet J.-L.;
RT "Genome evolution in yeasts."
RL Nature 430:35-44(2004).
DR ENBL; CR382128; CAG83062.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPRO11700; bZIP 2.
DR InterPro; IPRO04827; TF_bZIP.
DR Pfam; PF07716; bZIP 2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS0217; bZIP; 1.
DR PROSITE; PS00036; bZIP_BASIC; UNKNOWN 1.
KW Complete proteome; DNA-binding; Nuclear protein.
SQ COMPLETE 299 AA; 32817 MW; F526110CFB23ABB2 CRC64;

Query Match 54.3%; Score 175.5; DB 2; Length 299;
Best Local Similarity 63.8%; Pred. No. 6.1e-08;
Matches 37; Conservative 8; Mismatches 12; Indels 1; Gaps 1;

Qy 7 KRKSGQGVLPPEKTNLPKPKRAKTDEKEQRVERVLNRRAAQSRRKKLEVALE 64
|||
Db 33 KRKK-DLTPLPAGALPPKRAKTENEKEQRRIERIMNRQAHAHAREKKRHLEDLE 89
|||

RESULT 12
HACL1_YEAST STANDARD; PRT; 230 AA.
AC HACL1 YEAST
AC P41546; P87040;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
De HACL1 protein.
GN Name=HACL1; Synonyms=BRN4, IRE2; OrderedLocusNames=YFL031W;

Dungan J., Kuo A., Newport G., Lan C.-Y., Iijima C., Adegbola O.,
 RA Roberts J., Persson K., Donnelly S., Favoreto S., Tzung K.-W.,
 RA Jones T., Scherer S., Agabian N.;
 RA "Annotation of the Genome of *Candida albicans*.";
 RT Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AAC001000041; EAK9529.1; -; Genomic_DNA.
 DR EMBL; AAC001000040; EAK99617.1; -; Genomic_DNA.
 DR EMBL; AAC001000043; C:nucleus; IEA.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0003677; F:regulation of transcription, DNA-dependent; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 KW DNA-binding; Hypothetical protein; Nuclear protein.
 SQ SEQUENCE 357 AA; 40127 MW; 8E1C633A2B4ECD4E CRC64;

Query Match 53.9%; Score 174; DB 2; Length 357;
 Best Local Similarity 71.7%; Pred. No. 9.9e-08;
 Matches 33; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

Qy 19 KTNLPKPKRAKTEDEKEQRRVERVLNRRAAQSSRRERKRLVEALE 64
 : |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 54 KSTLPKPKRAKTEDEKEQRRVERVLNRRAAQSSRRERKRLVEALE 99
 : |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 14

Q75BQ5 ASHGQ
 ID Q75BQ5 ASHGQ PRELIMINARY; PRT; 228 AA.
 AC Q75BQ5;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE ACR216Cp.
 GN Name=ACR216C;
 OS *Ashbya gossypii* (Yeast) (*Eremothecium gossypii*).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Eremothecium.
 OX NCBI_TaxId=33169;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=ATCC 10895;
 RX PubMed=15001715; DOI=10.1126/science.1095781;
 RA Dietrich F.S., Voegelé S., Brachat S., Lerch A., Steiner S.,
 RA Mohr C., Pohlmann R., Luedi P., Choi S., Wing R.A., Flavier A.,
 RA Gaffney T.D., Philippen F.;
 RA "The *Ashbya gossypii* genome as a tool for mapping the ancient
 RT Saccharomycetes cerevisiae genome.";
 RL Science 304:304-307(2004).
 DR EMBL; AE016816; AAS51442.1; -; Genomic_DNA.
 DR AGD; ACR216C; -;
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR011700; bZIP 2.
 DR InterPro; IPR004827; TF bZIP.
 DR Pfam; PF07716; bZIP_2; 1.
 DR PROSITE; PS00036; bZIP_BASIC; 1.
 KW Complete proteome.
 SQ SEQUENCE 228 AA; 24914 MW; 605A60D78633F038 CRC64;

Query Match 52.0%; Score 168; DB 2; Length 228;
 Best Local Similarity 69.6%; Pred. No. 2.2e-07;
 Matches 32; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

Qy 19 KTNLPKPKRAKTEDEKEQRRVERVLNRRAAQSSRRERKRLVEALE 64
 : |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 14 KSTLPKPKRAKTEDEKEQRRVERVLNRRAAQSSRRERKRLVEALE 59
 : |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 15

Q6FLY3 CANGA
 ID Q6FLY3 CANGA PRELIMINARY; PRT; 329 AA.
 AC Q6FLY3;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:04:00 ; Search time 26.5576 Seconds
(without alignments)
1125.017 Million cell updates/sec

Title: US-10-663-450-60

Perfect score: 342

Sequence: 1 KSTLPKRAKTKKEQR.....CSLENNLNSVNEKLDHHE 68

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : 1_Geneseq_21.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	342	100.0	68	5 AAE15382	Aae15382 Yeast HAC
2	342	100.0	68	8 ADS12859	Adsi12859 Saccharom
3	342	100.0	84	5 ABP02534	Abp02534 Human ORF
4	342	100.0	200	8 ADS43437	Ads43437 Bacterial
5	342	100.0	230	2 AAW53806	Aaw53806 Transcrip
6	342	100.0	230	8 ADT87049	Adt87049 Yeast Str
7	342	100.0	238	2 AAW53807	Aaw53807 Transcrip
8	196	57.3	450	4 AAB82975	Aab82975 Trichoder
9	196	57.3	451	5 AAE15371	Aae15371 Trichoder
10	196	57.3	451	8 ADS12801	Adsi12801 Trichoder
11	196	57.3	451	8 ADS12804	Adsi12804 Trichoder
12	178.5	52.2	409	7 ABO43144	Abo43144 A. thalia
13	178.5	52.2	409	7 ADB31925	Adb31925 Plant (A.
14	178.5	52.2	409	8 ADO02271	Ado02271 Thalecres
15	176.5	51.6	342	4 AAB82977	Aab82977 Aspergill
16	176.5	51.6	342	5 AAE15381	Aae15381 Aspergill
17	176.5	51.6	342	8 ADS12818	Adsi12818 Aspergill
18	176.5	51.6	349	8 ADS12805	Adsi12805 Aspergill
19	176.5	51.6	349	8 ADS12803	Adsi12803 Aspergill
20	176.5	51.6	349	8 ADS12863	Adsi12863 Aspergill
21	176.5	51.6	350	4 AAB82976	Aab82976 Aspergill
22	176.5	51.6	350	5 AAE15372	Aae15372 Aspergill
23	176.5	51.6	386	5 AAE15379	Aae15379 Aspergill
24	176.5	51.6	386	8 ADS12815	Adsi12815 Aspergill

25	175	51.2	64	5 AAE15373	Aae15373 Trichoder
26	170	49.7	64	5 AAE15374	Aae15374 Aspergill
27	124.5	36.4	168	5 AAU93013	Aau93013 Arabidops
28	124.5	36.4	168	7 ADD30174	Add30174 Plant yie
29	124.5	36.4	168	8 ADI43893	Adi43893 Plant tra
30	124.5	36.4	211	3 AAG08861	Aag08861 Arabidops
31	121.5	35.5	192	4 AAB82614	Aab82614 Maize roo
32	121.5	35.5	192	4 AAB82615	Aab82615 Maize roo
33	121.5	35.5	192	4 AAB82616	Aab82616 Maize roo
34	121.5	35.5	192	4 AAG6525	Aag6525 Maize roo
35	121.5	35.5	192	4 AAG6526	Aag6526 Maize roo
36	116.5	34.1	170	8 ADM48147	Adm48147 Polypepti
37	115	33.6	23	3 AAY83623	Aay83623 Peptide f
38	114	33.3	185	9 ADM17162	Adm17162 Eucalyptu
39	112	32.7	163	9 ADWI7580	Adwi7580 Pinus rad
40	110	32.2	143	3 AAB33151	Aab33151 Pinus rad
41	107	31.3	120	3 AAG27808	Aag27808 Arabidops
42	107	31.3	135	3 AAG07181	Aag07181 Arabidops
43	107	31.3	149	3 AAG07180	Aag07180 Arabidops
44	107	31.3	149	3 AAG27807	Aag27807 Arabidops
45	107	31.3	188	3 AAG27806	Aag27806 Arabidops

ALIGNMENTS

RESULT 1

AAE15382

ID AAE15382 standard; protein; 68 AA.

AC AAE15382;

DT 07-MAR-2002 (first entry)

DE Yeast HAC1 protein DNA binding domain.

KW Heterologous protein secretion; unfolded protein response; UPR; lipase;
cellulase; carbohydrase; industry; purification; DNA binding domain;
HAC1 protein; Yeast.

OS Saccharomycetes cerevisiae.

PN US2001034045-A1.

PD 25-OCT-2001.

PF 23-MAR-2001; 2001US-00816277.

PR 24-MAR-2000; 2000US-00534692.

XX (GEMV) GENENCOR INT INC.

XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;

XX WPI; 2002-033728/04.

XX Increasing secretion of heterologous proteins e.g. lipase and cellulase
in eukaryotic cells useful in industry to increase production and
facilitate purification, by inducing an elevated unfolded protein
response.

XX Example 3; Fig 10; 56pp; English.

XX The present invention relates to methods for increasing the secretion of
heterologous protein in eukaryotic cells by inducing an elevated unfolded
protein response (UPR). The method involves inducing the elevated UPR by
increasing the presence of proteins such as HAC1, HACA, PTC2 or IRE1 in
cells. The method and sequences are useful for increasing the secretion
of heterologous proteins (e.g. lipase, cellulase, carbohydrase) in
eukaryotic cells useful in industry to increase protein yields and to
facilitate purification. The present sequence is yeast HAC1 protein DNA
binding domain

```
SQ Sequence 68 AA;
Query Match 100.0%; Score 342; DB 5; Length 68;
Best Local Similarity 100.0%; Pred. No. 2e-31;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KSTLPPRKAKTKKEEKEQRIERILNRRRAAHSREKRLHLQYLERKCSLLENLSYN 60
Db 1 KSTLPPRKAKTKKEEKEQRIERILNRRRAAHSREKRLHLQYLERKCSLLENLSYN 60
QY 61 LEKLADHE 68
Db 61 LEKLADHE 68
RESULT 2
ADS12859
ID ADS12859 standard; protein; 68 AA.
XX
AC ADS12859;
DT 16-DEC-2004 (first entry)
XX
DE Saccharomyces cerevisiae hac1 DNA binding domain.
XX
KW unfolded protein response; UPR; HAC1; PTC2;
KW unfolded protein response modulator; enzyme production; hac1; yeast;
KW chaperone; foldase; DNA binding domain.
XX
OS Saccharomyces cerevisiae.
XX
PN US2004186070-A1.
XX
PD 23-SEP-2004.
XX
PF 15-SEP-2003; 2003US-00663450.
XX
PR 24-MAR-2000; 2000US-00534692.
PR 23-MAR-2001; 2001US-00816277.
XX
XX (GEMV ) GENENCOR INT INC.
PA
PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
PI WPI; 2004-707924/69.
DR
XX
PT Increasing the secretion of a heterologous protein, such as a therapeutic
PT or an industrial enzyme, in genetically modified eukaryotic cells by
PT inducing an elevated unfolded protein response (UPR).
XX
PS Example 3; SEQ ID NO 60; 83pp; English.
XX
CC The invention describes a method of increasing the secretion of a
CC heterologous protein in a eukaryotic cell, comprising inducing an
CC elevated unfolded protein response (UPR). Also described are: an isolated
CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
CC unfolded protein response and has less than 50% similarity to yeast HAC1
CC protein, an isolated nucleic acid encoding a HAC1 protein, where the HAC1
CC protein induces unfolded protein response and where the HAC1 protein
CC comprises a DNA binding region that has greater than 70% similarity to
CC the DNA binding region of filamentous fungi HAC1 protein; a protein
CC having unfolded protein response inducing activity and having greater
CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
CC respectively) as given in the specification; a protein having an amino
CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
CC response and has at least 70% similarity to a fully defined amino acid
CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
CC specification; an isolated nucleic acid encoding a IRE1 protein that
CC modulates unfolded protein response and has at least 60% similarity to a
CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
CC having unfolded protein response modulating activity and having greater
```

```
than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
CC heterologous nucleic acid encoding a protein having unfolded protein
CC response modulating activity and a heterologous nucleic acid encoding a
CC protein of interest to be secreted. The methods and compositions of
CC genetically manipulating cells to have an elevated unfolded protein
CC response (UPR) resulting in an increased capacity to produce secreted
CC proteins, are useful in e.g. production of therapeutic or industrial
CC enzymes. This is the amino acid sequence of Saccharomyces cerevisiae hac1
CC chaperone and foldase DNA binding domain.
XX
SQ Sequence 68 AA;
Query Match 100.0%; Score 342; DB 8; Length 68;
Best Local Similarity 100.0%; Pred. No. 2e-31;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KSTLPPRKAKTKKEEKEQRIERILNRRRAAHSREKRLHLQYLERKCSLLENLSYN 60
Db 1 KSTLPPRKAKTKKEEKEQRIERILNRRRAAHSREKRLHLQYLERKCSLLENLSYN 60
QY 61 LEKLADHE 68
Db 61 LEKLADHE 68
RESULT 3
ABP02534
ID ABP02534 standard; protein; 84 AA.
XX
AC ABP02534;
XX
DT 24-JUN-2002 (first entry)
XX
DE Human ORFX protein sequence SEQ ID NO:5050.
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX
OS Homo sapiens.
XX
PN WO200192523-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US010836.
XX
PR 30-MAY-2000; 2000US-0206132P.
PR 29-AUG-2000; 2000US-0228716P.
XX
XX (CURA-) CURAGEN CORP.
XX
PI Shinketsu RA, Leach MD;
XX WPI; 2002-106308/14.
DR N-PSDB; ABN18286.
XX
PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders.
PS Disclosure; SEQ ID NO 5050; 1037pp; English.
XX
CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
```

CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage. N.B. The sequence data for this patent did not
 CC form part of the printed specification, but was obtained in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 84 AA;

Query Match 100.0%; Score 342; DB 5; Length 84;
 Best Local Similarity 100.0%; Pred. No. 2.5e-31;
 Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSTLPPKRAKTKKEKEQRIERILNRRRAAHQSREKKRLHLQYLERKCSLLENLLNSVN 60
 DB 15 KSTLPPKRAKTKKEKEQRIERILNRRRAAHQSREKKRLHLQYLERKCSLLENLLNSVN 74
 QY 61 LEKLADHE 68
 DB 75 LEKLADHE 82

RESULT 4
 ADS43437
 ID ADS43437 standard; protein; 200 AA.
 XX
 AC ADS43437;
 XX
 DT 02-DEC-2004 (first entry)
 XX
 DE Bacterial polypeptide #21867.

XX Recombinant DNA construct; transformed plant; improved plant property;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polypeptide.

OS Bacteria.

XX
 PN US2003233675-A1.
 XX
 PD 18-DEC-2003.
 XX
 PF 20-FEB-2003; 2003US-00369493.
 XX
 PR 21-FEB-2002; 2002US-0360039P.

XX (CAOV// CAO Y.
 PA (HINK// HINKLE G J.
 PA (SLAT// SLATER S C.
 PA (CHEN// CHEN X.
 PA (GOLD// GOLDMAN B S.

XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
 PI WPI; 2004-061375/06.
 XX
 DR
 XX

PT New recombinant DNA construct comprising a promoter positioned to provide
 PT for expression of a polynucleotide encoding a polypeptide from a
 PT microbial source, useful for producing plants with improved properties.

XX Claim 1; SEQ ID NO 21867; 122pp; English.

CC The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in a plant cell, where the promoter is positioned to
 CC provide for expression of a polynucleotide encoding a polypeptide from a
 CC microbial source. The invention also relates to a transformed plant
 CC comprising the recombinant DNA construct and a method of producing a
 CC transformed plant having an improved property. The plant is a crop plant
 CC such as maize or soybean. The method of producing a transformed plant
 CC having an improved property comprises transforming a plant with the
 CC recombinant DNA construct and growing the transformed plant, where the
 CC polynucleotide or polypeptide is useful for improving plant properties.
 CC The recombinant DNA construct is useful for producing plants with
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
 CC increased resistance to plant disease, better growth rate by modification
 CC of the cell cycle pathway with plant growth regulators, increased rate of
 CC homologous recombination, modified seed oil or protein yield and/or
 CC content, improved yield by modification of carbohydrate, nitrogen or
 CC phosphorus use and/or uptake, by modification of photosynthesis or by
 CC providing improved plant growth and development under at least one stress
 CC condition, improved lignin production or improved galactomannan
 CC production. This sequence represents a bacterial polypeptide used in the
 CC scope of the invention. Note: The sequence data for this patent did not
 CC form part of the printed specification but was obtained in electronic
 CC format from USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 200 AA;

Query Match 100.0%; Score 342; DB 8; Length 200;
 Best Local Similarity 100.0%; Pred. No. 6.3e-31;
 Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSTLPPKRAKTKKEKEQRIERILNRRRAAHQSREKKRLHLQYLERKCSLLENLLNSVN 60
 DB 23 KSTLPPKRAKTKKEKEQRIERILNRRRAAHQSREKKRLHLQYLERKCSLLENLLNSVN 82
 QY 61 LEKLADHE 68
 DB 83 LEKLADHE 90

RESULT 5
 AAW53806
 ID AAW53806 standard; protein; 230 AA.

XX AAW53806;

XX 06-JUL-1998 (first entry)

XX Transcription control protein ERN4.

XX Transcription control protein; ERN4; chaperone protein; protein folding;
 KW transcription enzyme.

XX Saccharomyces cerevisiae.

XX JP10084970-A.

XX 07-APR-1998.

XX 31-OCT-1996; 96JP-00307085.

XX 23-JUL-1996; 96JP-00213180.

XX (HSPK-) HSP KENKYUSHO KK.

XX WPI; 1998-264856/24.

XX N-PSDB; AAV20067.

XX DNA encoding a yeast transcriptional control protein - useful in the
PT recombinant production of high yield of protein of interest.
XX
PS Claim 1; Page 14-15; 21pp; Japanese.
XX
CC This sequence is the yeast transcriptional control protein, designated
CC ERN4, of the invention. An alternatively spliced version of the protein
CC is shown in AA52806. Yeast cells transformed with a vector containing
CC the DNA can be used for the recombinant production of a protein of
CC interest at a high yield. The DNA controls a chaperone protein which in
CC turn is responsible for the correct folding of an enzyme involved in
CC transcription, hence the cells can be used to control the transcription
CC of a recombinant protein of interest
XX
SQ Sequence 230 AA;

Query Match 100.0%; Score 342; DB 2; Length 230;
Best Local Similarity 100.0%; Pred. No. 7.3e-31;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSTLPKRAKTKKEQRIERILNRRAAHQSRKKRLHLOYLERKCSLLENLNSVN 60
DB 23 KSTLPKRAKTKKEQRIERILNRRAAHQSRKKRLHLOYLERKCSLLENLNSVN 82

QY 61 LEKLADHE 68
DB 83 LEKLADHE 90

RESULT 6
ADT87049
ID ADT87049 standard; protein; 230 AA.
XX
AC ADT87049;
XX
DT 13-JAN-2005 (first entry)
XX
XX Yeast Stress-related protein from gene YFL052W.
XX
XX Yeast; Stress-related protein; SRP; environmental stress; abiotic stress;
KW drought; heat; cold; salt.
XX
XX Saccharomyces cerevisiae.
XX
XX WO2004092398-A2.
XX
XX 28-OCT-2004.
XX
XX 15-APR-2004; 2004WO-US011888.
XX
XX 15-APR-2003; 2003EP-00008080.
PR 02-MAY-2003; 2003EP-00039728.
PR 01-AUG-2003; 2003EP-00016672.
PR 30-SEP-2003; 2003EP-00022225.
XX
XX (BADI) BASF PLANT SCI GMBH.
XX
XX Puzio P, Chardonnens A, Shirley A, Wang X, Sarria-Millan R;
PI McKersie B, Chen R;
PI
XX WPI; 2004-766883/75.
DR N-PSDB; ADT87048.
XX
XX New isolated nucleic acid molecule comprises a sequence encoding Stress-
PT Related Protein (SRP), useful for producing transformed plants with
PT altered metabolic activity resulting in increased tolerance or resistance
PT to environmental stress.
XX
XX Claim 15; SEQ ID NO 27; 91pp; English.
PS
XX The invention relates an isolated nucleic acid molecule comprises a
CC nucleic acid molecule encoding a Stress-Related Protein (SRP), from Yeast

CC or E. coli and their homologues from Rice, Soybean and Rape. Also
CC included are a transformed plant cell with altered metabolic activity
CC compared to a corresponding non-transformed wild type plant cell (where
CC the metabolic activity is altered by transformation with a SRP coding
CC nucleic acid and results in increased tolerance and/or resistance to an
CC environmental stress as compared to a corresponding non-transformed wild
CC type plant cell), a transgenic plant generated from the plant cell above
CC (and which is a monocot or dicot plant, or a gymnosperm plant), a seed
CC produced by a transgenic plant above (where the seed is genetically
CC homozygous for a transgene conferring altered metabolic activity
CC resulting in an increased tolerance to environmental stress as compared
CC to a corresponding non-transformed wild type plant), a nucleic acid
CC construct which confers the expression of the nucleic acid molecule above
CC (comprising one or more regulatory elements, where expression of the SRP
CC coding nucleic acid in a host cell results in altered metabolic activity
CC resulting in increased tolerance to environmental stress as compared to a
CC corresponding non-transformed wild type host cell), a vector comprising
CC the nucleic acid molecule above or the nucleic acid construct, a host
CC cell which has been transformed stably or transiently with the vector (or
CC the nucleic acid molecules above, or the nucleic acid construct), an
CC isolated Stress Related Protein (SRP) selected from the amino acid
CC sequences fully given in the specification and/or its homologues, a
CC method of producing a transgenic plant with altered metabolic activity
CC compared to a corresponding non-transformed wild type plant cell,
CC modifying stress tolerance of a plant, detecting environmental stress in
CC plant cells or plants, screening plant cells or plants for increased
CC tolerance and/or resistance to environmental stress, breeding plant cells
CC or plants towards increased tolerance and/or resistance to environmental
CC stress, increasing tolerance of a plant to at least one abiotic stress,
CC a plant transformed with the nucleic acids above and a seed of the plant.
CC The altered metabolic activity and/or a SRP encoding nucleic acids or its
CC homologues are useful as markers for selection of plants or plant cells
CC with increased tolerance to environmental stress, or for detection of
CC stress in plants or plant cells. The nucleic acids are useful for
CC producing transformed plants with altered metabolic activity resulting in
CC increased tolerance and/or resistance to an environmental stress
CC (drought, heat, cold and salt) as compared to a corresponding non-
CC transformed wild-type plant cell. The present sequence is a yeast SRP of
CC the invention.
XX
SQ Sequence 230 AA;

Query Match 100.0%; Score 342; DB 8; Length 230;
Best Local Similarity 100.0%; Pred. No. 7.3e-31;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSTLPKRAKTKKEQRIERILNRRAAHQSRKKRLHLOYLERKCSLLENLNSVN 60
DB 23 KSTLPKRAKTKKEQRIERILNRRAAHQSRKKRLHLOYLERKCSLLENLNSVN 82

QY 61 LEKLADHE 68
DB 83 LEKLADHE 90

RESULT 7
AAW53807
ID AAW53807 standard; protein; 238 AA.
XX
AC AAW53807;
XX
XX 06-JUL-1998 (first entry)
XX
XX Transcription control protein ERN4.
DE
XX Transcription control protein; ERN4; chaperone protein; protein folding;
KW transcription enzyme.
XX
XX Saccharomyces cerevisiae.
OS
XX JPI0084970-A.
PN
XX 07-APR-1998.
PD

```

XX PF 31-OCT-1996; 96JP-00307085.
XX XX
XX PR 23-JUL-1996; 96JP-00213180.
XX XX
XX PA (HSPK-) HSP KENYUSHO KK.
XX XX
XX DR WPI; 1998-264856/24.
XX DR N-PSDB; AAV20067.
XX XX
XX PT DNA encoding a yeast transcriptional control protein - useful in the
XX PT recombinant production of high yield of protein of interest.
XX PS Claim 2; Page 15; 21pp; Japanese.
XX XX
XX CC This sequence is the yeast transcriptional control protein, designated
XX CC ERN4, of the invention. An alternatively spliced version of the protein
XX CC is shown in AAW52805. Yeast cells transformed with a vector containing
XX CC the DNA can be used for the recombinant production of a protein of
XX CC interest at a high yield. The DNA controls a chaperone protein which in
XX CC turn is responsible for the correct folding of an enzyme involved in
XX CC transcription, hence the cells can be used to control the transcription
XX CC of a recombinant protein of interest
XX SQ Sequence 238 AA;
Query Match 100.0%; Score 342; DB 2; Length 238;
Best Local Similarity 100.0%; Pred. No. 7.6e-31;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KSTLPKRAKTEKEQRRIERILNRRRAAHQSRKRLHLQYLERKCSLLENLNSVN 60
DB 23 KSTLPKRAKTEKEQRRIERILNRRRAAHQSRKRLHLQYLERKCSLLENLNSVN 82
QY 61 LEKLADHE 68
DB 83 LEKLADHE 90
RESULT 8
AAB82975
ID AAB82975 standard; protein; 450 AA.
XX AC AAB82975;
XX DT 11-SEP-2003 (revised)
XX DT 21-DEC-2001 (first entry)
XX DE Trichoderma reesei HAC1, involved in unfolded protein response.
XX KW HAC1; transcription factor; unfolded protein response; protein secretion.
XX OS Hypocrea jecorina.
XX FH Key Location/Qualifiers
XX FT Domain 84..147
XX FT /label= DNA binding domain
XX PN WO200172783-A2.
XX PD 04-OCT-2001.
XX PF 23-MAR-2001; 2001WO-US009401.
XX PR 24-MAR-2000; 2000US-00534692.
XX PA (GENV ) GENENCOR INT INC.
XX PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX WPI; 2001-626252/72.
XX DR N-PSDB; AAW26931.
XX PT Increasing secretion of heterologous proteins e.g. lipase and cellulase

```

```

PT Increasing secretion of heterologous proteins e.g. lipase and cellulase
PT in eukaryotic cells useful in industry to increase production and
PT facilitate purification, by inducing an elevated unfolded protein
PT response.
XX PS Claim 54; Fig 7A-B; 89pp; English.
XX XX
XX CC The present sequence is that of the HAC1 protein of Trichoderma reesei,
XX CC as deduced from the newly isolated HAC1 gene (see AAW26931). HAC1 protein
XX CC is a transcription factor involved in the unfolded protein response
XX CC (UPR). The invention provides methods for increasing the secretion of a
XX CC heterologous protein in a cell by inducing an elevated UPR. This can be
XX CC achieved by modulating the activity of HAC1, PTC2 or IRE1 in the cell,
XX CC e.g. by gene overexpression. The cell from which the protein is secreted
XX CC can be any cell having an UPR, such as mammalian cells, insect cells,
XX CC yeast and filamentous fungi. The protein of interest can be any secreted
XX CC protein such as a therapeutic protein or an industrial enzyme, e.g.
XX CC lipase, cellulase, endoglucanase-H, protease, carboxylase, reductase,
XX CC oxidase, isomerase, transferase, kinase, phosphatase, alpha-amylase,
XX CC glucosylase, lignocellulose hemicellulase, pectinase and ligninase
XX CC (claimed). (Updated on 11-SEP-2003 to standardise OS field)
XX SQ Sequence 450 AA;
Query Match 57.3%; Score 196; DB 4; Length 450;
Best Local Similarity 61.9%; Pred. No. 7.5e-14;
Matches 39; Conservative 13; Mismatches 9; Indels 2; Gaps 1;
QY 1 KSTLPKRAKTEKEQRRIERILNRRRAAHQSRKRLHLQYLERKCSLLENLNSVN 60
DB 102 KTNLPKRAKTEKEQRRIERILNRRRAAHQSRKRLHLQYLERKCSLLENLNSVN 159
QY 61 LEK 63
DB 160 VQK 162
RESULT 9
AAE15371
ID AAE15371 standard; protein; 451 AA.
XX AC AAE15371;
XX DT 29-AUG-2003 (revised)
XX DT 07-MAR-2002 (first entry)
XX DE Trichoderma reesei HAC1 protein.
XX KW Heterologous protein secretion; unfolded protein response; UPR; lipase;
XX KW cellulase; carboxylase; industry; purification; HAC1 protein.
XX OS Hypocrea jecorina.
XX FH Key Location/Qualifiers
XX FT Binding-site 84..147
XX FT /label= DNA-binding_domain
XX PN US2001034045-A1.
XX PD 25-OCT-2001.
XX PF 23-MAR-2001; 2001US-00816277.
XX PR 24-MAR-2000; 2000US-00534692.
XX PA (GENV ) GENENCOR INT INC.
XX PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX WPI; 2002-033728/04.
XX DR N-PSDB; AAD24595.
XX PT Increasing secretion of heterologous proteins e.g. lipase and cellulase

```


ID ADB31925 standard; protein; 409 AA.
 XX ADB31925;
 AC
 XX
 XX
 DT 04-DEC-2003 (first entry)
 DE Plant (A. thaliana) transcription factor polypeptide #79.
 DE
 XX
 KW Plant; transcription factor; transgenic plant; transgenic; plant trait;
 KW modified trait.
 KW
 XX Arabidopsis thaliana.
 OS
 XX
 PN US2003101481-A1.
 XX
 XX
 PD 29-MAY-2003.
 XX
 XX
 PF 15-NOV-2002; 2002US-00295403.
 XX
 XX
 PR 22-SEP-1998; 98US-0101349P.
 PR 06-OCT-1998; 98US-0103312P.
 PR 17-NOV-1998; 98US-0108734P.
 PR 22-DEC-1998; 98US-0113409P.
 PR 13-SEP-1999; 99US-00394519.
 XX
 XX (ZHAN/) ZHANG J.
 XX (FROM/) FROMM M.
 XX (HEAR/) HEARD J.
 XX (RIEC/) RIECHMANN J L.
 XX (ADAM/) ADAM L.
 XX (BROU/) BROUN P.
 XX (PINE/) PINEDA O.
 XX (REUB/) REUBER L.
 XX (KEDD/) KEDDIE J.
 XX (YUGG/) YU G.
 XX (JIAN/) JIANG C.
 XX
 XX Zhang J, Fromm M, Heard J, Riechmann JL, Adam L, Broun P;
 PI Pineda O, Reuber L, Keddie J, Yu G, Jiang C;
 XX
 DR WPI; 2003-597572/56.
 DR N-PSDB; ADB31924.
 XX
 XX New isolated polypeptides and polynucleotide sequences, useful for
 PT screening a molecule to identify a molecule that modifies plant trait,
 PT and for producing plants with modified traits.
 XX
 XX Claim 9; SEQ ID NO 158; 17pp; English.
 XX
 CC The present invention relates to the isolation of plant (Arabidopsis
 CC thaliana) transcription factor polypeptide and polynucleotide sequences.
 CC Also disclosed are: an expression vector comprising the isolated
 CC polynucleotide, a host cell comprising the expression vector, a
 CC transgenic plant comprising the isolated polynucleotide, a transgenic
 CC plant ectopically expressing the isolated polynucleotide or polypeptide,
 CC a method for screening a molecule to identify a molecule that modifies a
 CC plant trait by placing the molecule in contact with the plant, and
 CC monitoring the effect of the molecule on the expressing or activity of
 CC the polypeptide or polynucleotide, and producing a transgenic plant
 CC having a modified trait by ectopically expressing the isolated
 CC polypeptide and selecting a plant with the modified trait. The
 CC polypeptides, polynucleotides and methods are useful for screening a
 CC molecule to identify a molecule that modifies plant trait, and for
 CC producing plants with modified traits. The present sequence represents a
 CC plant transcription factor polypeptide of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site at seqdata.uspto.gov.
 XX
 XX Sequence 409 AA;
 SQ
 Query Match 52.2%; Score 178.5; DB 7; Length 409;
 Best Local Similarity 56.5%; Pred. No. 6.8e-12;

Matches 39; Conservative 14; Mismatches 13; Indels 3; Gaps 2;
 Qy 1 KSTLPPPKRAKTEKEQRRIERILNRRAAHOSEKRLHLQYLERKCSLLENLNSVN 60
 Db 84 KTNLPPPKRAKTDQEKQRRVERVLNRRAAQSRERKRQVEALEVEKRAIER--KNMD 141
 Qy 61 LE-KLADHE 68
 Db 142 LEMRLADME 150
 RESULT 14
 ADO02271
 ID ADO02271 standard; protein; 409 AA.
 XX
 XX ADO02271;
 AC
 XX
 XX
 DT 01-JUL-2004 (first entry)
 DE
 XX
 XX Thalecress transcription factor protein #342.
 XX
 KW Thalecress; transcription factor; plant; transgenic; abiotic stress;
 KW cold tolerance; heat tolerance; drought; osmotic stress;
 KW phosphate limitation; potassium limitation; nitrogen limitation;
 KW hormone sensitivity; disease resistance; sugar sensing; seed germination;
 KW flowering; inflorescence architectural change;
 KW meristem cell differentiation; phyllotaxy; apical dominance;
 KW trichome development; seed development; premature senescence;
 KW delayed senescence; lethality; necrosis; plant size; leaf morphology;
 KW seed morphology; secondary metabolism; light response; shade avoidance.
 XX
 XX Arabidopsis thaliana.
 OS
 XX
 XX
 PN US2004045049-A1.
 XX
 XX
 PD 04-MAR-2004.
 XX
 XX
 PF 10-APR-2003; 2003US-00412699.
 XX
 XX 13-SEP-1999; 99US-00394519.
 PR 21-JAN-2000; 2000US-00489376.
 PR 17-FEB-2000; 2000US-00506720.
 PR 22-MAR-2000; 2000US-00532591.
 PR 22-MAR-2000; 2000US-00533029.
 PR 22-MAR-2000; 2000US-00533030.
 PR 22-MAR-2000; 2000US-00533392.
 PR 22-MAR-2000; 2000US-00533648.
 PR 06-APR-2000; 2000MO-US009448.
 PR 16-NOV-2000; 2000US-00713994.
 PR 17-MAR-2001; 2001US-00819142.
 PR 17-APR-2001; 2001US-00837444.
 PR 30-JAN-2002; 2002US-00958131.
 PR 14-JUN-2002; 2002US-00171468.
 PR 09-AUG-2002; 2002US-00225066.
 PR 09-AUG-2002; 2002US-00225067.
 PR 09-AUG-2002; 2002US-00225068.
 PR 17-DEC-2002; 2002US-0434166P.
 PR 25-FEB-2003; 2003US-00374780.
 XX (ZHAN/) ZHANG J.
 XX (FROM/) FROMM M E.
 XX (HEAR/) HEARD J E.
 XX (RIEC/) RIECHMANN J L.
 XX (ADAM/) ADAM L J.
 XX (BROU/) BROUN P E.
 XX (PINE/) PINEDA O.
 XX (REUB/) REUBER T L.
 XX (KEDD/) KEDDIE J S.
 XX (YUGG/) YU G.
 XX (JIAN/) JIANG C.
 XX (SAMA/) SAMAHA R S.
 XX (PILG/) PILGRIM M L.
 XX (CREE/) CREELMAN R A.

Matches 38; Conservative 12; Mismatches 17; Indels 1; Gaps 1;

Qy 1 KSTLPPRKRAKTKEEKEQRRIERILRNRRAAHQSRKKRLHLQYLERKCSLLENLLNSVN 60

Db 64 KTNLPPRKRAKTEDEKEQRRIERVLRNFNFAAQTSSRRKRLEMEKLENEKIQMEQ-QNQFL 122

Qy 61 LEKLADHE 68

Db 123 LQRLSQME 130

Search completed: November 23, 2005, 03:18:41

Job time : 27.5576 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:18:52 ; Search time 4.42281 Seconds
(without alignments)
239.171 Million cell updates/sec

Title: US-10-663-450-4

Perfect score: 1766

Sequence: 1 MKSADRFSPVKMEDAFANSP.....PSHGASTSRCGQGIAGSA 349

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 17545 seqs, 3030971 residues

Total number of hits satisfying chosen parameters: 17545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New:

- 1: /cgn2_5/ptodata/2/pubpaa/US10_NEW_PUB_PEP.*
- 2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB_PEP.*
- 3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB_PEP.*
- 4: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB_PEP.*
- 5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB_PEP.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB_PEP.*
- 7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB_PEP.*
- 8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB_PEP.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	192	10.9	376	1	US-10-971-994-2
2	131.5	7.4	431	1	US-10-821-234-1285
3	121.5	6.9	605	1	US-10-131-826A-160
4	121	6.9	588	7	US-11-022-562-210
5	118.5	6.7	1618	1	US-10-984-645-2
6	113	6.4	1138	1	US-10-509-422-4
7	104.5	5.9	1168	1	US-10-509-422-2
8	102.5	5.8	1616	1	US-10-821-234-1497
9	102	5.8	298	1	US-10-821-234-1240
10	101	5.7	4384	1	US-10-821-234-1120
11	100.5	5.7	2432	1	US-10-821-234-899
12	98	5.5	431	7	US-11-169-013-2
13	98	5.5	454	7	US-11-169-013-1
14	97.5	5.5	216	1	US-10-821-234-925
15	97	5.5	865	7	US-11-109-156-4
16	97	5.5	1122	1	US-10-821-234-1657
17	95	5.4	401	1	US-10-821-234-881
18	95	5.4	1410	1	US-10-821-234-1050
19	94	5.3	676	7	US-11-135-855-28
20	94	5.3	717	7	US-11-135-855-29
21	93.5	5.3	413	1	US-10-967-648A-8
22	92	5.2	777	1	US-10-821-234-1658
23	91.5	5.2	135	1	US-10-821-234-1037
24	91.5	5.2	1377	1	US-10-821-234-1070
25	91	5.2	1142	7	US-11-044-051-73

26	90.5	5.1	1637	1	US-10-821-234-1204	Sequence 1204, Ap
27	90	5.1	1207	1	US-10-821-234-1109	Sequence 1109, Ap
28	89	5.0	1189	7	US-11-074-176-134	Sequence 134, App
29	89	5.0	1560	7	US-11-059-982-1	Sequence 1, Appli
30	88.5	5.0	655	1	US-10-793-626-1052	Sequence 1052, Ap
31	88.5	5.0	655	1	US-10-793-626-1400	Sequence 1400, Ap
32	88.5	5.0	915	1	US-10-821-234-1514	Sequence 1514, Ap
33	88	5.0	1155	1	US-10-793-626-1780	Sequence 1780, Ap
34	86.5	4.9	240	1	US-10-689-742-210	Sequence 210, App
35	86.5	4.9	373	7	US-11-082-389-258	Sequence 258, App
36	86.5	4.9	400	7	US-11-018-018-7	Sequence 7, Appli
37	86.5	4.9	400	7	US-11-047-757-7	Sequence 16, Appl
38	86.5	4.9	691	1	US-10-131-826A-16	Sequence 963, App
39	86.5	4.9	703	1	US-10-821-234-963	Sequence 252, App
40	86.5	4.9	761	1	US-10-485-517-252	Sequence 1294, Ap
41	86	4.9	445	1	US-10-793-626-1294	Sequence 862, App
42	85.5	4.8	290	1	US-10-821-234-862	Sequence 2, Appli
43	85.5	4.8	437	1	US-10-967-648A-2	Sequence 12, Appl
44	85.5	4.8	677	1	US-10-982-545-12	Sequence 132, App
45	85	4.8	387	1	US-10-485-517-132	

ALIGNMENTS

RESULT 1

US-10-971-994-2
; Sequence 2, Application US/10971994
; Publication No. US20050250182A1
; GENERAL INFORMATION:
; APPLICANT: University of Michigan et al.
; APPLICANT: Kaufman, Randal
; APPLICANT: Kyungo, Lee
; APPLICANT: Mori, Kazutos
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, AND METHODS FOR
; TITLE OF INVENTION: MODULATING THE UNFOLDED PROTEIN RESPONSE
; FILE REFERENCE: UMV-2246
; CURRENT APPLICATION NUMBER: US/10/971,994
; CURRENT FILING DATE: 2004-10-21
; PRIOR APPLICATION NUMBER: PCT/2003/012640
; PRIOR FILING DATE: 2003-04-22
; PRIOR APPLICATION NUMBER: 60/375,098
; PRIOR FILING DATE: 2002-04-22
; PRIOR APPLICATION NUMBER: 60/374,880
; PRIOR FILING DATE: 2002-04-23
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-971-994-2

Query Match 10.9%; Score 192; DB 1; Length 376;

Best Local Similarity 26.0%; Pred. No. 2.6e-08;
Matches 93; Conservative 52; Mismatches 127; Indels 86; Gaps 17;

QY	33	SPADTSLRKNVVAQPKPEKKPAKKRKGQELP--VP-----KTNLP-PRKRA 79
DB	9	NPADGTPKVLISGGQPASAAAPA-----GQALPLMVPAQRGASPEASGGLPQARKQ 62
QY	80	K-TEDEKQRRIVRLNRNRAAQAQTSRKRLEMEKLESEKIDMEQOQFL-----QR 131
DB	63	RUTLSPEKALRRKLKNRVAQAQTRDKKARMSLEQQVQVQVLEENQKLLLENQKLLREK 122
QY	132	LAQMEANNRLSQO-----VAQLSAEVGTS--RHSTPTSSSPASVSPITPTLFLKQEGD 183
DB	123	THGLVVENQELQRLGMDALVAEEAEAKGNEVRPVAGSAEAGAGPVVTPP-----E 176
QY	184	EYPLDR-----IPPTPSVTDVSPTLKPSLSLAESPDLTQHHPAVSVGGLE 227
DB	177	HLPMDGGIDSSDSDDLGLDNLDPVNFPCPSPEPASLEELFEV--YP-----E 227

```
QY 228 GDES-----ALTFLDGLASIKHEPTHDTAPLSDDDFRRLFNGDSSLESDDSLLEDGFAPFV 284
Db 228 GPSSLPASLSLVSQTSKALEINELIR-----FDHIYTKPLVLEIPSETESQANVVVK 281
QY 285 LSGDLSAPFDS---WVDFTEPVTLEDEQTNGLSDSASCKASLQPSHGASTSRC 339
Db 282 IEAPLSPSNDHPHFIVSKBPVE-DDLVPELGISN-----LLSSHCPCPKSSC 331

RESULT 2
US-10-821-234-1285
; Sequence 1285, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1285
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1285

Query Match 7.4%; Score 131.5; DB 1; Length 431;
Best Local Similarity 21.7%; Pred. No. 0.0014;
Matches 80; Conservative 47; Mismatches 129; Indels 113; Gaps 15;

QY 4 ADRF-----SPVKMEDAFANSTTPPSLEVPLTVSPADTSLRTKNVVVAQTKPEKPAKR 59
Db 85 ADSFSMGSFVNQAQDFCTDLAVSSANFIP--TVTAISTSPDLQWLQPALVSSVAPSQTR 142
QY 60 KSWGQBLVPKTNLP--KRAKTE-----DEKQRIERVLNR 99
Db 143 APHFPGVPAPFAGSAGYRAGVVKMTGGAQSIGRGKVEQLSPDEEKRIIRRE-RNKMA 201
QY 100 AQTSRERKLEMEKLESEKIDMEQNFLLQRLAQMEANNRLS----- 143
Db 202 AAKCNRRLDTLQAEQDQLEKSAQTEIANLLKEKLEFETLAHRPACKIPDDL 261
QY 144 ---QQVAQLSAEVRGSRHSTPTSSSPASVPTLTPTLFKQEGDEVPLDRIPPTPSVTDY 200
Db 262 GPPEMSVASLDLTG---GLPEVATPES-----EEAFTPLPNDPEKPSV--- 304
QY 201 SPTLKPSLAESPDLTQHFAVSGVGLGEGDESALTFLDGLASIKHEPTHDTAPLSDDDR 260
Db 305 -----EPVKSISSME-----LKTEPFDLFPASSRP--- 331
QY 261 RLFPNGDSSLES--DSSLLEDGFAPD---VLDSDLSAPFDSWVDFTEPVTLEDEQTN 316
Db 332 ---SGSETARSPVDMDLSGSFYAADWEPLHSGSLGMPMAT-----BLEPLC-----TPV 378
QY 317 LSDASCKA 325
Db 379 VTCPTPCTA 387

RESULT 3
US-10-131-826A-160
; Sequence 160, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
```

```
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 160
; LENGTH: 605
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-131-826A-160

Query Match 6.9%; Score 121.5; DB 1; Length 605;
Best Local Similarity 23.4%; Pred. No. 0.013;
Matches 82; Conservative 45; Mismatches 107; Indels 117; Gaps 19;

QY 52 EKKPAKRSWCOEL-----PVPKTNL-----PPRKAETDEKEORRIERVLR 99
Db 125 EKAGSIETQSQOELFPLNPLSPKPNLVPEPPWHMPPREEEEEEER----- 173
QY 100 AQTSRERKLEMEKLESEKIDMEQNFLLQRLAQMEANNRLSQQVAQL-----SAEVRG 155
Db 174 -----EKEVEKQEEEEELLPVNGSQEAK-----PQVRDFSLTSSSQTPG 216
QY 156 ---SRH---STPTSSSPASVPTLTPTLFKQEGDEVPLDRIPPTPSVTDYSPTLKP---S 207
Db 217 ATKSRHSDSGDQASSGVEVESSMGPSLL-----LPSVTPT-----TVTGDQD 259
QY 208 SLAESPDLTQHFAVSVG-----GLEGDESALTFLDGLASIKHEPTHDL-----TAPLS- 255
Db 260 STSQAEATVLPAAGLGVFEAPQASEEA-TAGAGLSGQHEEVPALSPFOPTTAPSGA 318
QY 256 ---DDD-----FRRLFNGDSSLESDDL-EDGFAPDVLSDGLSAPFDSWVDFDT 303
Db 319 EHPDEDPLGSRTSASSPLAPGMELTPSSATLGQEDLNQQLLEG---QAAEQSRIPWDS 375
QY 304 EPTVLEDEQTN-----LSDSASCK-----AASLPQSHGA 334
```


	Matches	75; Conservative	63; Mismatches	163; Indels	60; Gaps	14;
Qy	17	ANSPPTPSLEVPVLTVSPADTSURTKNVVAQTPEKKP-AKRRKSGOELPV---	BKTN	72		
Dd	373	ANSTAATS---SVLTIQSSATPVK---VPAPGFESNHKPKGALRPGNGSEVLMVYGPPQQ	426			
Qy	73	LPRPKRAKTEDEKEORRIERVLNRAAAQTSRRKRLEMEKLESEKIDMEQONQFLLQRL	132			
Dd	427	PPOQHVRLOQLQGDMWLQGLHLHRPHHHHQOQOQOQOQOQOQOQOQOQOQOQLQNA	486			
Qy	133	AQAEAKNNRLSQOVAQLSAEVRGSRHSTPTSSSPASVSPTLTP-----TLFKQ	180			
Dd	487	YLOQYCHAMHQHILQOQFLM----HSVYOQPQPASQYPAMMQYQAFLQQQMLARHQ	542			
Qy	181	EGDEVPLDRIPFP---TPSVDTSPTLKPSLSAESPDLTQHPAVSVGGLEGDESALTFD	237			
Dd	543	PAQQVSPEYLTSPOEPSPALVSYASSL-PAQVGITVDVSSYGANRSV---AEKEAVANFT	597			
Qy	238	LGASIKHEPTHDLTAPLSDDDFRLLFGDSSLSLSDSSILLDEGFADFVLDGSDLASA-PFFD	296			
Dd	598	NQKTI SHPPDMSGWNPFGEDNFSKL-----TEEELLDR--EFDLLRNRLGASTPFD	647			
Qy	297	SMWDFOTEPVTL---EDLEQTNGLSDSASCKAASLOPS-----HGASTSRCDG	341			
Dd	648	KTVDLPPAHSRPPEEPFASVPFIHSGSPSEKKTTHSPNQKSITANLTKNNGSSPLCKD	707			
Qy	342	Q	342			
Dd	708	Q	708			

```

RESULT 7
US-10-509-422-2
; Sequence 2, Application US/10509422
; Publication No. US20050244825A1
; GENERAL INFORMATION:
; APPLICANT: Liou, Simon
; TITLE OF INVENTION: Human BMP2 Inducible Kinases
; FILE REFERENCE: 004974.01015
; CURRENT APPLICATION NUMBER: US/10/509,422
; CURRENT FILING DATE: 2004-09-24
; PRIOR APPLICATION NUMBER: PCT/EP03/080825
; PRIOR FILING DATE: 2003-03-20
; PRIOR APPLICATION NUMBER: US 60/367,512
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/406,936
; PRIOR FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1168
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-509-422-2

```

```

Db      549  -----MLAQHQBSQQQAPEYLTSQEFSPALVSVTSSLPQAQVGTIMDSSYANRQVF 601
Qy      227  ---EGDESALTFLDGLGASIKHEPTHDTLAPLSDDDFRRLFNGLSDSSLESLLLEDGFAPD 283
Db      602  FQSVADKEAIAFTNQKNISNPPDMGWNPFGEIDNFSKL-----TEBELLDR--EFD 651
Qy      284  VLDSDGL 290
Db      652  LLRSNRL 658

RESULT 8
US-10-821-234-1497
; Sequence 1497, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821.234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1497
; LENGTH: 1616
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1497

```

RESULT 9
US-10-821-234-1240
; Sequence 1240, Application US/10821234
; Publication No. US2005025511A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan


```
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preclampsia
; FILE REFERENCE: 821A
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1240
; LENGTH: 298
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1240

Query Match      5.8%; Score 102; DB 1; Length 298;
Best Local Similarity 27.0%; Pred. No. 0.16;
Matches 41; Conservative 18; Mismatches 49; Indels 44; Gaps 6;

QY 15 AFANSPPTPSLEVVLTVSPADTSRLTKNVVAQTPPEKKPAKRSWGQELPVPKTNLP 74
Db 179 AAAGOPTPTSPPEPRSS-SPROT-----PAPGPAREKSAGKRG-----PDGSGP 221
QY 75 PKRAKTEDEKQRIERVLNRNAAQTSRERKLEMEKLESEKIDMEQOQFLQRLAQ 134
Db 222 -----EYQRR-----ERNNIAVRKSRDKAK-----RRNQEMQOKLVE 254

QY 135 MEAENNRLSQQVAQLSAEVRGSRHSTPTSSSP 166
Db 255 LSAENEKHLHQRVEQLTRDLAGLQRFQFKQLPSP 286

RESULT 10
US-10-821-234-1120
; Sequence 1120, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preclampsia
; FILE REFERENCE: 821A
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1120
; LENGTH: 4384
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1120

Query Match      5.7%; Score 101; DB 1; Length 4384;
Best Local Similarity 21.9%; Pred. No. 7.4;
Matches 73; Conservative 54; Mismatches 139; Indels 68; Gaps 16;

QY 5 DRFSPVKMEDAFANSPPTPSLEVVL-TVSPADTSRLTKNVVAQTPPEKKPAKRSWG 63
Db 3158 DTLEQVSLDSGSGKSPLTP--ETPSSESVSEFTSKTPDLSIAYIP-----G 3202
QY 64 QELVPKTN--LPPRKRAKTEDEK---EORRIERVLNRNAAQTSRERKLEMEK----- 111
Db 3203 KPSPIPEVSESEEEQAKSTLSKQTTVEETFAVEREMENDVSKNQRPKNRVAYIEFP 3262
QY 112 --EKLESEKIDMEQOQFLQRLAQMEAENNRLSQQVAQLSAEVRGSRHSTPTSSSP-AS 168
Db 3263 PPPPLDADQIESDKKHHLPEKEVDWIEVNLQDEHDKQLAEPV--IRVQPPSPVPPGAD 3320
```

```
QY 169 VSPTLTPTLTKQEGDEVPLDRIPFPPTPSVTDYSPTLKPSSLAESPDLTQHPAVSVGLEG 228
Db 3321 VSDSSDDESIYQ---PVPVKYTFKLKEVDDQKE-KPKASAEKAS-NQKELESNGSGKD 3375
QY 229 DESALTLFDLGASIKHEPTHDLTAPLSDDDFRLLFNGDSSLSLSDSLLEDGFAFDVLDG 288
Db 3376 NE-----FGLGLD---SQNEIAQ-----NGN-----NDQSITCSATTAEPFH 3412
QY 289 DLSAPFPDSMVDFTPEPVTLEDLEQTNGLSDSAS 322
Db 3413 DTDATEIDSLDGYDLQ-----DSDGGLTESDS 3439

RESULT 11
US-10-821-234-899
; Sequence 899, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 899
; LENGTH: 2432
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-899

Query Match      5.7%; Score 100.5; DB 1; Length 2432;
Best Local Similarity 21.9%; Pred. No. 3.6;
Matches 86; Conservative 53; Mismatches 147; Indels 107; Gaps 16;

QY 14 DAFANSPPTPSLEV-----PVLTVSPADTSRLTKN-----VVAQTPPEKKPAK 57
Db 54 DAASASRLPNEEIVQKIEEVLG-VLDTLRYKPDLEKSGRSKRCVSVQTDPTDEIPTK 112
QY 58 KRKSWGQELPVPKTNLPPRKRAKTEDEKEQRIERVLNRNAAQTSRERKLEMEK----- 113
Db 113 KSKK-----HKHKNNKKKKKEKKYKRQPESESKT---KSHDDGNIDLESDFSPL 162
QY 114 -----LSEKIDMEQOQFLQRL--AQMEAENNR 141
Db 163 KFDSEPSAVALLELPTRAFGPSETNESPAVVLEPPVVGMEVSEPHILETLKPAKTAELSV 222
QY 142 LSQOVAQLSAEVRGSRHSTPT-----SSSPASVSPTLTPTLTKQEGDEVPLDRIPFPPTS 196
Db 223 VTSVISEQSESVAVMPEPSMTKILDSFAAAPVPTTLVLKSE-----PVVTMS 273
QY 197 VTDYSPTLKPSSLAESPD-----LTQHPAVSVGLEGDESALTFLDLGASIKHEPTHOLT 251
Db 274 VEYQMKSVLKSVESTSPSPSKIMLVEPPVAKV--LEPSETLVWSSETPTVVPPEPSTST 331
QY 252 APLSDDDFRLLFNGDSSLSLSDSLLEDGFAFDVLDG--DLSAFPFDSDMVDFTPEPVTLED 310
Db 332 M-----DF-----PESSAIE---AURLPEQVPDVPSEIADSSM---TRQELPE 369
QY 311 LEQTNGLSDSASCKAASLQPSHGASTSRCDGQG 343
Db 370 LPKTTALEQESSVASAMELPGPATSMPELOG 402

RESULT 12
US-11-169-013-2
; Sequence 2, Application US/11169013
; Publication No. US20050244971A1
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:12:56 ; Search time 112.993 Seconds
(without alignments)
1290.661 Million cell updates/sec

Title: US-10-663-450-4
Perfect score: 1766
Sequence: 1 MKSADRFSPVKMEDAFANSP.....PSHGASTSRCDGGIAGA 349

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA Main:
1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pap:*
2: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pap:*
3: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pap:*
4: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pap:*
5: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pap:*
6: /cgn2_6/prodata/1/pubpaa/US11_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1766	100.0	349	3	US-09-816-277-4
2	1766	100.0	349	3	US-09-816-277-6
3	1766	100.0	349	4	US-10-663-450-4
4	1766	100.0	349	4	US-10-663-450-6
5	1293.5	73.2	342	3	US-09-816-277-19
6	1293.5	73.2	342	4	US-10-663-450-19
7	1293.5	73.2	386	3	US-09-816-277-16
8	1293.5	73.2	386	4	US-10-663-450-16
9	434	24.6	451	3	US-09-816-277-2
10	434	24.6	451	3	US-09-816-277-5
11	434	24.6	451	4	US-10-663-450-2
12	434	24.6	451	4	US-10-663-450-5
13	398.5	22.6	409	3	US-09-533-029-104
14	398.5	22.6	409	4	US-10-295-403-158
15	398.5	22.6	409	4	US-10-412-699B-684
16	318	18.0	174	4	US-10-767-701-61238
17	216	12.2	200	4	US-10-369-493-21867
18	176.5	10.0	68	3	US-09-816-277-60
19	176.5	10.0	68	4	US-10-663-450-60
20	167.5	9.5	203	4	US-10-437-963-151695
21	167.5	9.5	203	5	US-10-732-923-13694
22	167.5	9.5	203	5	US-10-732-923-13695
23	164.5	9.3	176	4	US-10-437-963-164243
24	164.5	9.3	634	5	US-10-741-849-7083
25	164	9.3	370	5	US-10-864-348-2
26	163.5	9.3	168	3	US-09-934-455-102
27	163.5	9.3	168	4	US-10-225-066A-206

28	163.5	9.3	168	4	US-10-374-780A-2356	Sequence 2356, Ap
29	163.5	9.3	168	5	US-10-732-923-13692	Sequence 13692, A
30	163.5	9.3	168	5	US-10-225-066A-206	Sequence 206, App
31	160.5	9.1	147	4	US-10-425-115-187302	Sequence 187302,
32	160	9.1	261	4	US-10-205-823-453	Sequence 453, App
33	160	9.1	261	4	US-10-177-293-504	Sequence 504, App
34	160	9.1	261	4	US-10-755-889-763	Sequence 763, App
35	160	9.1	261	6	US-11-051-454-453	Sequence 453, App
36	160	9.1	261	6	US-11-095-893-12	Sequence 12, Appl
37	160	9.1	264	4	US-10-264-049-2772	Sequence 2772, Ap
38	159.5	9.0	260	4	US-10-177-293-502	Sequence 502, App
39	159.5	9.0	371	6	US-11-095-893-4	Sequence 4, Appli
40	158	8.9	721	4	US-10-286-264-42	Sequence 42, Appl
41	158	8.9	721	4	US-10-412-699B-368	Sequence 368, App
42	155.5	8.8	192	3	US-09-772-656-10	Sequence 10, Appl
43	155.5	8.8	192	4	US-10-396-199A-10	Sequence 10, Appl
44	155	8.8	360	4	US-10-437-963-165206	Sequence 165206,
45	155	8.8	369	4	US-10-425-115-314993	Sequence 314993,

ALIGNMENTS

RESULT 1

US-09-816-277-4
; Sequence 4, Application US/09816277
; Publication NO. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-816-277-4

Query Match		100.0%;	Score 1766;	DB 3;	Length 349;
Best Local Similarity		100.0%;	Pred. No. 3.2e-118;	Mismatches 0;	Indels 0;
Matches 349;		Conservative 0;			
Qy	1	MKSADRFSPVKMEDAFANSPPTTSPLEVPVLTSPADTSLRTKNVAAQTSPKPKAKRK	60		
Db	1	MKSADRFSPVKMEDAFANSPPTTSPLEVPVLTSPADTSLRTKNVAAQTSPKPKAKRK	60		
Qy	61	SWGQLPVPKTNLPPRKAKTDEKEQRRIERVLNRAAAQTSPKPKAKRK	120		
Db	61	SWGQLPVPKTNLPPRKAKTDEKEQRRIERVLNRAAAQTSPKPKAKRK	120		
Qy	121	MEQQOFLQRLAQAENNRSLQVAQSAEVRGSRHSPTSSSPASVSPTLTPTLFQK	180		
Db	121	MEQQOFLQRLAQAENNRSLQVAQSAEVRGSRHSPTSSSPASVSPTLTPTLFQK	180		
Qy	181	EGDEVPDLRIPTTSPVTDYSPTLKPSLSLAESPDLTQHPAVSGVGLGDESALTFLDGA	240		
Db	181	EGDEVPDLRIPTTSPVTDYSPTLKPSLSLAESPDLTQHPAVSGVGLGDESALTFLDGA	240		
Qy	241	SIKHEPHTDLTAPLSDDDFRRLFNQDSSLESDSSLLDGGFAFDVLDGDLSPFDSMVD	300		
Db	241	SIKHEPHTDLTAPLSDDDFRRLFNQDSSLESDSSLLDGGFAFDVLDGDLSPFDSMVD	300		
Qy	301	FTEPVTLEDLEQTNGLSASCKASLQPSHGASTSRCDGGIAGA	349		

Db 301 FDTPEVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDGGGIAAGSA 349
|||||

RESULT 2
US-09-816-277-6
; Sequence 6, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816, 277
; CURRENT FILING DATE: 2001-03-23
; PRIOR FILING DATE: US 09/534, 692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-816-277-6

Query Match 100.0%; Score 1766; DB 3; Length 349;
Best Local Similarity 100.0%; Pred. No. 3.2e-118;
Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKSADRFSPVKMEDAFANSPPTPSLEVPVLTVPADTSLRTKNVVAQTPEEKKPAKKRK 60
Db 1 MKSADRFSPVKMEDAFANSPPTPSLEVPVLTVPADTSLRTKNVVAQTPEEKKPAKKRK 60
QY 61 SWGOELPVPKTNLPKRKRAKTEDEKEQRIERVLNRNRAAQTSLRTKNVVAQTPEEKKLESEKID 120
Db 61 SWGOELPVPKTNLPKRKRAKTEDEKEQRIERVLNRNRAAQTSLRTKNVVAQTPEEKKLESEKID 120
QY 121 MEQONQFLQLRLAQMEAEENRLSQVAQLSAEVRGSRHSTPTSSSPASVSPTLTPTLTKQ 180
Db 121 MEQONQFLQLRLAQMEAEENRLSQVAQLSAEVRGSRHSTPTSSSPASVSPTLTPTLTKQ 180
QY 181 EGDEVPLDRIPPTPSVTDYSPTLKPSLSAESPDLTQHPAVSVGGLEGDESALTFLDLGA 240
Db 181 EGDEVPLDRIPPTPSVTDYSPTLKPSLSAESPDLTQHPAVSVGGLEGDESALTFLDLGA 240
QY 241 SIKHEPTHDLTAPLSDDDFRRLFNGLDSSLSLESDDLGDGFAFDVLDGSLSAFFPDSMVD 300
Db 241 SIKHEPTHDLTAPLSDDDFRRLFNGLDSSLSLESDDLGDGFAFDVLDGSLSAFFPDSMVD 300
QY 301 FDTPEVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDGGGIAAGSA 349
Db 301 FDTPEVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDGGGIAAGSA 349

RESULT 3
US-10-663-450-4
; Sequence 4, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663, 450
; CURRENT FILING DATE: 2003-09-15

; PRIOR APPLICATION NUMBER: US 09/534, 692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-663-450-4
Query Match 100.0%; Score 1766; DB 4; Length 349;
Best Local Similarity 100.0%; Pred. No. 3.2e-118;
Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKSADRFSPVKMEDAFANSPPTPSLEVPVLTVPADTSLRTKNVVAQTPEEKKPAKKRK 60
Db 1 MKSADRFSPVKMEDAFANSPPTPSLEVPVLTVPADTSLRTKNVVAQTPEEKKPAKKRK 60
QY 61 SWGOELPVPKTNLPKRKRAKTEDEKEQRIERVLNRNRAAQTSLRTKNVVAQTPEEKKLESEKID 120
Db 61 SWGOELPVPKTNLPKRKRAKTEDEKEQRIERVLNRNRAAQTSLRTKNVVAQTPEEKKLESEKID 120
QY 121 MEQONQFLQLRLAQMEAEENRLSQVAQLSAEVRGSRHSTPTSSSPASVSPTLTPTLTKQ 180
Db 121 MEQONQFLQLRLAQMEAEENRLSQVAQLSAEVRGSRHSTPTSSSPASVSPTLTPTLTKQ 180
QY 181 EGDEVPLDRIPPTPSVTDYSPTLKPSLSAESPDLTQHPAVSVGGLEGDESALTFLDLGA 240
Db 181 EGDEVPLDRIPPTPSVTDYSPTLKPSLSAESPDLTQHPAVSVGGLEGDESALTFLDLGA 240
QY 241 SIKHEPTHDLTAPLSDDDFRRLFNGLDSSLSLESDDLGDGFAFDVLDGSLSAFFPDSMVD 300
Db 241 SIKHEPTHDLTAPLSDDDFRRLFNGLDSSLSLESDDLGDGFAFDVLDGSLSAFFPDSMVD 300
QY 301 FDTPEVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDGGGIAAGSA 349
Db 301 FDTPEVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDGGGIAAGSA 349

RESULT 4
US-10-663-450-6
; Sequence 6, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663, 450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534, 692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-663-450-6

Query Match 100.0%; Score 1766; DB 4; Length 349;
Best Local Similarity 100.0%; Pred. No. 3.2e-118;
Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKSADRFSPVKMEDAFANSPPTPSLEVPVLTVPADTSLRTKNVVAQTPEEKKPAKKRK 60
Db 1 MKSADRFSPVKMEDAFANSPPTPSLEVPVLTVPADTSLRTKNVVAQTPEEKKPAKKRK 60
QY 61 SWGOELPVPKTNLPKRKRAKTEDEKEQRIERVLNRNRAAQTSLRTKNVVAQTPEEKKLESEKID 120

```
Db 61 SWGQLPVKTNLPKRAKTEDEKEORRIERVLRNRAAAQTSRERKLEMEKLESKID 120
Qy 121 MEQONQFLQRLAQMEANNRLSQVAQLSAEVRGSRHSTPTSSSPASVSPTLTPTLFKQ 180
Db 121 MEQONQFLQRLAQMEANNRLSQVAQLSAEVRGSRHSTPTSSSPASVSPTLTPTLFKQ 180
Qy 181 EGEVPLDRIPFPPTPSVTDYSPTLKPSSLAESPDLTQHPAVSVGGLEGDESALTFLDLGA 240
Db 181 EGEVPLDRIPFPPTPSVTDYSPTLKPSSLAESPDLTQHPAVSVGGLEGDESALTFLDLGA 240
Qy 241 SIKHEPTHDTAPLSDDDPRLFNGLSDSSLSLLEDGFAFDVLDGDLSPAFPDMSVD 300
Db 241 SIKHEPTHDTAPLSDDDPRLFNGLSDSSLSLLEDGFAFDVLDGDLSPAFPDMSVD 300
Qy 301 FDTSPVLTLEDQTNGLSDSASCAASLQPSHGASTSRCDGQGIAGSA 349
Db 301 FDTSPVLTLEDQTNGLSDSASCAASLQPSHGASTSRCDGQGIAGSA 349
```

RESULT 5

US-09-816-277-19

; Sequence 19, Application US/09816277

; Publication No. US20010034045A1

; GENERAL INFORMATION:

; APPLICANT: Penttila, Merja E.

; APPLICANT: Ward, Michael

; APPLICANT: Wang, Huaming

; APPLICANT: Valkonen, Mari J.

; APPLICANT: Saloheimo, Markku

; TITLE OF INVENTION: Increased Production of Secreted

; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells

; FILE REFERENCE: GC590-2

; CURRENT APPLICATION NUMBER: US/09/816,277

; PRIOR FILING DATE: 2001-03-23

; PRIOR FILING DATE: 2000-03-24

; NUMBER OF SEQ ID NOS: 63

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 19

; LENGTH: 342

; TYPE: PRT

; ORGANISM: Aspergillus niger

US-09-816-277-19

```
Query Match 73.2%; Score 1293.5; DB 3; Length 342;
Best Local Similarity 76.2%; Pred. No. 2.1e-84;
Matches 262; Conservative 29; Mismatches 46; Indels 7; Gaps 3;

Qy 5 DRFSPVKMEDAFANSPPTPSLEVPVLTSPADTSLRTKNVA-OTKPEKKPAKPKKSWG 63
Db 4 EAFSPV---DSLGSPTP---ELPLLTSPADTSLDSSVQAGETKAEKKPKPKKSWG 57
Qy 64 QELPVKTNLPKRAKTEDEKEORRIERVLRNRAAAQTSRERKLEMEKLESEKIDMEQ 123
Db 58 QELPVKTNLPKRAKTEDEKEORRIERVLRNRAAAQTSRERKLEMEKLESEKIDMEQ 117
Qy 124 QNQLLQRLAQMEANNRLSQVAQLSAEVRGSRHSTPTSSSPASVSPTLTPTLFKQEGD 183
Db 118 QNQLLQRLAQMEANNRLSQVAQLSAEVRGSRGNTPKGPSVPSASPTLTPTLFKQERD 177
Qy 184 EYPLDRIPFPPTPSVTDYSPTLKPSSLAESPDLTQHPAVSVGGLEGDESALTFLDLGASIK 243
Db 178 EYPLDRIPFPPTPSVTDYSPTLKPSSLAESPDLTQHPAVSVGGLEGDESALTFLDLGASIK 237
Qy 244 HEPTHDTAPLSDDDPRLFNGLSDSSLSLLEDGFAFDVLDGDLSPAFPDMSVD 303
Db 238 PHAADLAPLSDDDPRLFNGLSDSSLSLLEDGFAFDVLDGDLSPAFPDMSVD 297
Qy 304 EYPTLEDQTNGLSDSASCAASLQPSHGASTSRCDGQGIAG 347
Db 298 EYVGFEGIEPPHGLPDETSTRQTSVQPSLGASTRCDGQGIAG 341
```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-09-816-277-16

Query Match
Best Local Similarity 73.2%; Score 1293.5; DB 3; Length 386;
Matches 262; Conservative 29; Mismatches 46; Indels 7; Gaps 3;

Qy 5 DRFSPVKMEDAFANSPPTTSLEVPVLTVSPADTSLRTKNVA-QTKPEEKPAKRRKSWG 63
Db 48 EAFSPV---DSLGSPTP---ELPLTVSPADTSLDSSVQAGETKAEKKPVKRRKSWG 101
Qy 64 QELPVPKTNLPPRKRAKTEDEKEORRIERVLRNRAAQTSTRERKELEMEKLESEKIDMEQ 123
Db 102 QELPVPKTNLPPRKRAKTEDEKEORRIERVLRNRAAQTSTRERKELEMEKLESEKIQMEQ 161
Qy 124 QNQFLLQRLQSAEANNRLSQVAQLSAEVRGSRHSTPTSSSPASVSTLTPTTLFKQEGD 183
Db 162 QNQFLLQRLQSAEANNRLNQVAQLSAEVRGSRGNTPKPGSPVSASTLTPTTLFKQERD 221
Qy 184 EVELDRIPPTPSVTDYSTLKPSSLAESPDLTQHPAVSVGGLRGDSALTLPDLGASIK 243
Db 222 EIPLERIPPTPSITDYSTLRPSTLAESSDVTQHPAVSVAGLGEAGSALSLFDVGSNPE 281
Qy 244 HEPTHTDLTAPLSDDDFRRLFNQDSSLESDDSLLEDGFAFDVLDGDLGASAFDDSMVDFDT 303
Db 282 PHAADDLAAPLSDDDFHLFNVDSVPGSDSSVLEDDGFAFDVLDGDLGASAFDDSMVDFDP 341
Qy 304 EPTVLEDTQNLGSLDASCKAASLQPSHGASTSRCDCQGIAAG 347
Db 342 ESVGFGIEPPHGLPDETSTRQTSVQPSLGASTSRCDCQGIAAG 385

RESULT 8
US-10-663-450-16
; Sequence 16, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; PRIOR FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-10-663-450-16

Query Match
Best Local Similarity 73.2%; Score 1293.5; DB 4; Length 386;
Matches 262; Conservative 29; Mismatches 46; Indels 7; Gaps 3;

Qy 5 DRFSPVKMEDAFANSPPTTSLEVPVLTVSPADTSLRTKNVA-QTKPEEKPAKRRKSWG 63
Db 48 EAFSPV---DSLGSPTP---ELPLTVSPADTSLDSSVQAGETKAEKKPVKRRKSWG 101
Qy 64 QELPVPKTNLPPRKRAKTEDEKEORRIERVLRNRAAQTSTRERKELEMEKLESEKIDMEQ 123
Db 102 QELPVPKTNLPPRKRAKTEDEKEORRIERVLRNRAAQTSTRERKELEMEKLESEKIQMEQ 161
Qy 124 QNQFLLQRLQSAEANNRLSQVAQLSAEVRGSRHSTPTSSSPASVSTLTPTTLFKQEGD 183
Db 162 QNQFLLQRLQSAEANNRLNQVAQLSAEVRGSRGNTPKPGSPVSASTLTPTTLFKQERD 221
Qy 184 EVELDRIPPTPSVTDYSTLKPSSLAESPDLTQHPAVSVGGLRGDSALTLPDLGASIK 243
Db 222 EIPLERIPPTPSITDYSTLRPSTLAESSDVTQHPAVSVAGLGEAGSALSLFDVGSNPE 281
Qy 244 HEPTHTDLTAPLSDDDFRRLFNQDSSLESDDSLLEDGFAFDVLDGDLGASAFDDSMVDFDT 303
Db 282 PHAADDLAAPLSDDDFHLFNVDSVPGSDSSVLEDDGFAFDVLDGDLGASAFDDSMVDFDP 341
Qy 304 EPTVLEDTQNLGSLDASCKAASLQPSHGASTSRCDCQGIAAG 347
Db 342 ESVGFGIEPPHGLPDETSTRQTSVQPSLGASTSRCDCQGIAAG 385

RESULT 9
US-09-816-277-2
; Sequence 2, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-09-816-277-2

Query Match
Best Local Similarity 24.6%; Score 434; DB 3; Length 451;
Matches 149; Conservative 50; Mismatches 109; Indels 152; Gaps 21;

Qy 5 DRFSPVKMEDAFANSPPTTSLEVPVLTVSPADTSLRTKNVAQ----- 47
Db 26 DNFTSL-----PADS--TPS-----TLNPRD--MMPDSDVADISRLSVIPESQDAED 70
Qy 48 -----TKP--EEKKPAKRRKSGQELPVPKTNLPPRKRAKTEDEKEORRIERVLRNRA 98
Db 71 ESHSTSATAPSTSEKPKVKKRSGQVLPPEKTNLPPRKRAKTEDEKEORRIERVLRNRR 130
Qy 99 AAQTSRERKRLEMEKLESEK-----IDMEQQQFLQLRLQAQMEANNRLS----- 143
Db 131 AAQSSRERKRLEVEALEKRKKELETLINQVQTNLILVEELNRRFRSSGVVTRSSPLDS 190
Qy 144 -QOVAQLSAEVRGSRH-----STPTSS-----SPASVSTLTPTTLFKQ----- 180
Db 191 LQDSITLQQQLFGSRDGTMTNSPEQSLMDQIMRGAANPTVNPASLSPSLPISDKBQTK 250
Qy 181 EGDEVPDLRIPPPTPSVTDYSTLKPSSLA-----SPDLTQHPAVSVGGLGDESA 232
Db 251 EDEFSQAE---DEMEQTHETKEAATAAEEKSKQSRVSTDTQRPAVSTGG-----DAA 303
Qy 233 LTLF--DLGAS--IKHEPHTDLTAPLSDDDFRRLFNQDSSLESDDSLLEDGFAFDVLDGSD 289
Db 304 VPFVSDDAGANCLGLDPVHQDDGPPS---IGHSGLSAALDADRYLLE-----SQ 350
Qy 290 LSAPFPDSDMVDP-----TEPVTLF--DLQTNGLSDSASCKAASL----- 328
Db 351 LLAASNAVSTVDDYLAGDSAAACFTNPPLSPSDYDFINDFLTDDANHAAYDIVAASNAAYAD 410

```


QY 329 -----QPSHGASTRCDGGGIAAG 347
DB 411 RELDLEIHDPENQIPSRHSIQPQSGASHGCDGGIAGV 450

RESULT 10

US-09-816-277-5
; Sequence 5, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Wang, Michael
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-09-816-277-5

Query Match 24.6%; Score 434; DB 3; Length 451;
Best Local Similarity 32.4%; Pred. No. 1e-22;
Matches 149; Conservative 50; Mismatches 109; Indels 152; Gaps 21;

QY 5 DRFSPVKMEDAFANSPPTPSLEVPVLTVPADTSLRTKNVQA-----47
DB 26 DNFTSL-----FADS--TPS-----TLNPRD--MMTPDSVADISRLSVIPESQDAEDD 70
QY 48 -----TKP--EKKPAKRRKSWGQELPVPKTNLPKRAKTEDEKEQRRVERVLNR 98
DB 71 ESHSTSATAPSTSEKKPVKRRKSWGQVLPKTNLPKRAKTEDEKEQRRVERVLNR 130
QY 99 AAOQSRERKRLMEKLESEK-----IDMEQOQFLQRLQMAEANNRLS-----143
DB 131 AAQSSRRERKRLVEALEKRNKELETLINVOKTNLILVELNRRFRSSGVVTRSSPLDS 190
QY 144 -QOVAQLSAEVRGSRH---STPTSS-----SPASVSPILTPTLFQK---180
DB 191 LQDSITLSQQLFGSRDQGTMSNPEQSLMDQIMRSAAANTVNPASLSPSPISDKKEFQTK 250
QY 181 EGDEVPLDRIPPTPSVTDSYPTLKPSLAE-----SPDLTQHPAVSVGGLGDESA 232
DB 251 EDEEQADE--DEEMEQTWHTKEAAAKENKSKQSRVSTDTORPAVSIIG---DAA 303
QY 233 LTLF--DLGAS--IKHEPTHDLTAPLSDDDFRRLFNQDSSLESDDLDFGFAFVLDSDG 289
DB 304 VPVFSDDAGANCLGLDPVHQDDGPPS---IGHSFGLSAAALDADRYLLE-----SQ 350
QY 290 LSAPFPDSMVDFD-----TEPVTLE--DLEQTNGLSDSASCKAASL-----328
DB 351 LLASPNASTVDDDYLAGDSAACTNPLPSDYDFDINFLTDDANHAAYDIVAASNYAAAD 410
QY 329 -----QPSHGASTRCDGGGIAAG 347
DB 411 RELDLEIHDPENQIPSRHSIQPQSGASHGCDGGIAGV 450

RESULT 11

US-10-663-450-2
; Sequence 2, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:

; APPLICANT: Penttila, Merja E.
; APPLICANT: Wang, Michael
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-663-450-2

Query Match 24.6%; Score 434; DB 4; Length 451;
Best Local Similarity 32.4%; Pred. No. 1e-22;
Matches 149; Conservative 50; Mismatches 109; Indels 152; Gaps 21;

QY 5 DRFSPVKMEDAFANSPPTPSLEVPVLTVPADTSLRTKNVQA-----47
DB 26 DNFTSL-----FADS--TPS-----TLNPRD--MMTPDSVADISRLSVIPESQDAEDD 70
QY 48 -----TKP--EKKPAKRRKSWGQELPVPKTNLPKRAKTEDEKEQRRVERVLNR 98
DB 71 ESHSTSATAPSTSEKKPVKRRKSWGQVLPKTNLPKRAKTEDEKEQRRVERVLNR 130
QY 99 AAOQSRERKRLMEKLESEK-----IDMEQOQFLQRLQMAEANNRLS-----143
DB 131 AAQSSRRERKRLVEALEKRNKELETLINVOKTNLILVELNRRFRSSGVVTRSSPLDS 190
QY 144 -QOVAQLSAEVRGSRH---STPTSS-----SPASVSPILTPTLFQK---180
DB 191 LQDSITLSQQLFGSRDQGTMSNPEQSLMDQIMRSAAANTVNPASLSPSPISDKKEFQTK 250
QY 181 EGDEVPLDRIPPTPSVTDSYPTLKPSLAE-----SPDLTQHPAVSVGGLGDESA 232
DB 251 EDEEQADE--DEEMEQTWHTKEAAAKENKSKQSRVSTDTORPAVSIIG---DAA 303
QY 233 LTLF--DLGAS--IKHEPTHDLTAPLSDDDFRRLFNQDSSLESDDLDFGFAFVLDSDG 289
DB 304 VPVFSDDAGANCLGLDPVHQDDGPPS---IGHSFGLSAAALDADRYLLE-----SQ 350
QY 290 LSAPFPDSMVDFD-----TEPVTLE--DLEQTNGLSDSASCKAASL-----328
DB 351 LLASPNASTVDDDYLAGDSAACTNPLPSDYDFDINFLTDDANHAAYDIVAASNYAAAD 410
QY 329 -----QPSHGASTRCDGGGIAAG 347
DB 411 RELDLEIHDPENQIPSRHSIQPQSGASHGCDGGIAGV 450

RESULT 12

US-10-663-450-5
; Sequence 5, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Wang, Michael
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692

QY 10 VKMEDAFANSPPTPSLEVPVLTVSPA--DTSLRTKNVVAQTKPEEKPKAKRKSQGOELP 67
Db 22 VSTPSSFHNPPLFDNNLNPVDGFSPOSFDRDYNFNGSLGLNLPKPKPKRKSQGOQLP 81
QY 68 VPKTNLPKRKRAKTEDEKEORRIERVLRNRAAAQTSRERKRLEMEKLESEKIDMEQONQF 127
Db 82 EPKTNLPKRKRAKTQDEKEQRRVLRNRAAQSRRERKQVEALEVEKRAIERKNMD 141
QY 128 LLQRLAQMEANNRLSQOVAQLSAE-----VRGSRHSTPTSSPASVSP-TLTPTLFKOE 181
Db 142 LEMRLADMEAKYLLQQLKRAASGYNKNFLSYSDSTPDISDSQSLPITFSKQLFNAQ 201
QY 182 GDEVPDLDRIP---FPTPSVTDYSPTLKPSSL-----AESDPLTOHPAVSVGGLE 227
Db 202 -DELCPISQSIGPLTSRTVDPSTLSPKSLSPDSSNSNSDMDTQHPAVVLCDLQ 256

RESULT 15

US-10-412-699B-684
; Sequence 684, Application US/10412699B
; Publication No. US20040045049A1
; GENERAL INFORMATION:
; APPLICANT: Mendel Biotechnology, Inc.
; APPLICANT: Zhang, James
; APPLICANT: Fromm, Michael E.
; APPLICANT: Heard, Jacqueline E.
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Adam, Luc J.
; APPLICANT: Broun, Pierre E.
; APPLICANT: Pineda, Omaira
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Keddle, James S.
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Samaha, Raymond R.
; APPLICANT: Pilgrim, Marsha L.
; APPLICANT: Creelman, Robert A.
; APPLICANT: DuBell, Arnold N.
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Kumimoto, Roderick
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: Polynucleotides and Polypeptides in Plants
; FILE REFERENCE: MBI-0048CIP
; CURRENT APPLICATION NUMBER: US/10/412,699B
; CURRENT FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: 09/394,519
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 09/489,376
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: 09/506,720
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 09/533,030
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,392
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,029
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/532,591
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,648
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/713,994
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: 09/819,142
; PRIOR FILING DATE: 2001-03-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2011
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 684
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:

; OTHER INFORMATION: G1034
US-10-412-699B-684

Query Match 22.6%; Score 398.5; DB 4; Length 409;
Best Local Similarity 45.3%; Pred. No. 3.2e-20;
Matches 107, Conservative 26; Mismatches 84; Indels 19; Gaps 6;
QY 10 VKMEDAFANSPPTPSLEVPVLTVSPA--DTSLRTKNVVAQTKPEEKPKAKRKSQGOELP 67
Db 22 VSTPSSFHNPPLFDNNLNPVDGFSPOSFDRDYNFNGSLGLNLPKPKPKRKSQGOQLP 81
QY 68 VPKTNLPKRKRAKTEDEKEORRIERVLRNRAAAQTSRERKRLEMEKLESEKIDMEQONQF 127
Db 82 EPKTNLPKRKRAKTQDEKEQRRVLRNRAAQSRRERKQVEALEVEKRAIERKNMD 141
QY 128 LLQRLAQMEANNRLSQOVAQLSAE-----VRGSRHSTPTSSPASVSP-TLTPTLFKOE 181
Db 142 LEMRLADMEAKYLLQQLKRAASGYNKNFLSYSDSTPDISDSQSLPITFSKQLFNAQ 201
QY 182 GDEVPDLDRIP---FPTPSVTDYSPTLKPSSL-----AESDPLTOHPAVSVGGLE 227
Db 202 -DELCPISQSIGPLTSRTVDPSTLSPKSLSPDSSNSNSDMDTQHPAVVLCDLQ 256

Search completed: November 23, 2005, 03:32:21
Job time : 114.983 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM protein - protein search, using sw model

Run on: November 23, 2005, 03:12:51 ; Search time 34.1763 Seconds
(without alignments)
844.265 Million cell updates/sec

Title: US-10-663-450-4
Perfect score: 1766
Sequence: 1 MKSADRFSPVKMEDAFANSP.....PSHGASTSRCDGGIAGSA 349

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: /cgn2_6/ptodata/1/iaa/5 COMB.pcp.*
 - 2: /cgn2_6/ptodata/1/iaa/6 COMB.pcp.*
 - 3: /cgn2_6/ptodata/1/iaa/H COMB.pcp.*
 - 4: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pcp.*
 - 5: /cgn2_6/ptodata/1/iaa/RE COMB.pcp.*
 - 6: /cgn2_6/ptodata/1/iaa/backfile01.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

				SUMMARIES	
Result No.	Score	Query Match	Length DB ID	Description	
1	398.5	22.6	409 2	US-09-533-029-104	Sequence 104, App
2	194	11.0	248 2	US-09-248-796A-18840	Sequence 18840, A
3	160	9.1	264 2	US-09-949-016-10789	Sequence 10789, A
4	159.5	9.0	260 2	US-09-538-092-950	Sequence 950, App
5	145.5	8.2	1402 2	US-09-248-796A-14503	Sequence 14503, A
6	141.5	8.0	521 1	US-08-721-684C-2	Sequence 2, Appli
7	141.5	8.0	521 1	US-09-005-970-2	Sequence 2, Appli
8	141.5	8.0	521 2	US-09-407-715-2	Sequence 2, Appli
9	141	8.0	340 2	US-09-350-841A-1601	Sequence 1601, Ap
10	137.5	7.8	143 2	US-09-840-211A-1930	Sequence 1930, Ap
11	135.5	7.7	395 2	US-09-247-155-113	Sequence 113, App
12	135.5	7.7	395 2	US-09-513-999C-14	Sequence 14, Appl
13	135.5	7.7	395 2	US-09-471-276-14	Sequence 14, Appl
14	135.5	7.7	395 2	US-09-903-190-113	Sequence 113, App
15	134.5	7.6	380 2	US-09-165-522-16	Sequence 16, Appl
16	133	7.5	102 2	US-09-840-211A-1046	Sequence 1046, Ap
17	133	7.5	338 1	US-08-318-686-2	Sequence 2, Appli
18	133	7.5	338 2	US-08-460-242-2	Sequence 2, Appli
19	133	7.5	1341 2	US-09-949-016-6890	Sequence 6890, Ap
20	133	7.5	1344 2	US-09-949-016-10925	Sequence 10925, A
21	132.5	7.5	380 2	US-08-486-099-115	Sequence 115, App
22	132.5	7.5	380 2	US-08-484-223B-115	Sequence 115, App
23	132.5	7.5	380 2	US-08-919-597-115	Sequence 115, App
24	132.5	7.5	380 2	US-08-475-668A-115	Sequence 115, App
25	132.5	7.5	380 2	US-08-485-551A-115	Sequence 115, App
26	132.5	7.5	380 2	US-08-471-913A-115	Sequence 115, App
27	132.5	7.5	380 2	US-08-485-264A-115	Sequence 115, App

Sequence 115, App
Sequence 115, App
Sequence 115, App
Sequence 115, App
Sequence 115, App
Sequence 24, Appl
Sequence 24, Appl
Sequence 24, Appl
Sequence 125, App
Sequence 7289, Ap
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 22, Appl
Sequence 22, Appl
Sequence 14492, A
Sequence 9676, Ap

ALIGNMENTS

RESULT 1
US-09-533-029-104
; Sequence 104, Application US/09533029
; Patent No. 6664446
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Broun, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Keddle, James
; APPLICANT: Pineda, Omaira
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-liang
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Reuber, Lynne
; TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES
; FILE REFERENCE: MBI-010
; CURRENT APPLICATION NUMBER: US/09/533,029
; EARLIER FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 60/125,814
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 104
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1034
US-09-533-029-104

Query Match 22.6%; Score 398.5; DB 2; Length 409;
Best Local Similarity 45.3%; Pred. No. 8.8e-29;
Matches 107; Conservative 26; Mismatches 84; Indels 19; Gaps 6;
QY 10 VKMEDAFANSPPTPSLEVPVLTVSPA--DTSURTKNVVAQTPEEKPAKRSWQQLP 67
DB 22 VSTPSFTHNPPLFDNLNPNVDGFSQSFDRYFNFGSLSLGLNLPEKKPKKRSWQQLP 81
QY 68 VPKTNLPKRKAKTDEKEORIERVLRNRAAQTSSRERKLEMEKLESEKIDMEQNOF 127
DB 82 EPKTNLPKRKAKTQDEKEQRVERVLRNRAAQTSSRERKLEMEKLESEKIDMEQNOF 141
QY 128 LIQRLAOMEAENNRSLQQVAQLSAE-----VRGSRHSTPTSSSPASVSP-TLPTTLFKOE 181
DB 142 LEMRLADMEAKYLLQQLKRAAGYKNTFLSYSDSSTPDISDSQLSPITFSKQLFNAQ 201

QY 182 GDEVPLDRI-FTFSPVTDYSPILKPSL-----AESDLTQHPAVSVGGLE 227
Db 202 -DELCPISQSIGLTSITVDPESTLSPKSLSPDSSNSNSDMDTQHPAVVLCDLQ 256

RESULT 2
US-09-248-796A-18840
; Sequence 18840, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18840
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Candida albicans
; NAME/KEY: UNSURE
; LOCATION: (221)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unkno
US-09-248-796A-18840

Query Match 11.0%; Score 194; DB 2; Length 248;
Best Local Similarity 30.1%; Pred. No. 5.7e-10;
Matches 59; Conservative 33; Mismatches 60; Indels 44; Gaps 6;
QY 11 KMEDAFANSPPTPSLEVPVLTSPADTSLRTKNVVAQTKPEEKPAKRRKSWGQELPVP- 69
Db 18 KMELTVDNNTTNSID-----DLSVATPTSLMTSTTTSPSMSTSTSSHNTLIDIP 68
QY 70 ---KTNLPKRAKTEDEKQRIERVLNRNRAAQTSSRRKRLMEKLESEKIDME---- 122
Db 69 ATFKTLPKRAKTEDEKQRIERVLNRNRAAQTSSRRKRLMEKLESEKIDME---- 128
QY 123 -----QONQFLQRLQAMEAENRLSQVAQLSAEVRGSRHSTPTSSSPAS 168
Db 129 KUNNNYNQAFELTKDNQELL--LSKLEVLDD-VSDLKEQIHNSMGTRRRSHNKSNDED 185
QY 169 VSPTLTPTLTKQEGDE 184
Db 186 I-----EEDE 191

RESULT 3
US-09-949-016-10789
; Sequence 10789, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10789
; LENGTH: 264

; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10789
Query Match 9.1%; Score 160; DB 2; Length 264;
Best Local Similarity 30.3%; Pred. No. 9.6e-07;
Matches 56; Conservative 29; Mismatches 66; Indels 34; Gaps 8;
QY 33 SPADTSLRTKNVVAQTKPEEKPAKRRKSWGQELP--VP-----KTNLP-PRKEA 79
Db 12 NPADGTPKVLISGQFASAGAPA-----GQALPLMVPQAGSPAAASGGLPQARKQ 65
QY 80 K-TEDEKQRIERVLNRNRAAQTSSRRKRLMEKLESEKIDMEQONQFLL-----QR 131
Db 66 RLTHLSPEKALRRKLKNRVAQATDRKKARMSLEQQQVVDLEENQKLLLENQLLREK 125
QY 132 LAQMEAEENRLSQ-----VAQLSAEVRGSRHSTPTSSSPASVSPTLTPTLTKQEGDEV 185
Db 126 THGLVVENQELRQLGMDALVAEEBAEAKGN-EVRPVAGSAESAALRLRAPLQQVOAQLS 184
QY 186 PLDRI 190
Db 185 PLQNI 189
RESULT 4
US-09-538-092-950
; Sequence 950, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurapSeqFormat Version 0.9
; SEQ ID NO 950
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P17861
US-09-538-092-950

Query Match 9.0%; Score 159.5; DB 2; Length 260;
Best Local Similarity 28.8%; Pred. No. 1e-06;
Matches 55; Conservative 31; Mismatches 80; Indels 25; Gaps 6;
QY 15 AFANSPPTPSLEVPVLTSPADTSLRTKNVVAQTKPEEKPAKRRKSWGQELPVPKTNLP 74
Db 5 AAAPNPADGTPKVLISGQFASAGAPA-----AAGAPAAKPLMVPQAGSPAAASGGLP 55
QY 75 -PRKEAK-TEDEKQRIERVLNRNRAAQTSSRRKRLMEKLESEKIDMEQONQFLL--- 129
Db 56 QARKEQLTHLSPEKALRRKLKNRVAQATDRKKARMSLEQQQVVDLEENQKLLLEN 115
QY 130 ---ORLAQMEAEENRLSQ-----VAQLSAEVRGSRHSTPTSSSPASVSPTLTPTLTK 179
Db 116 QLLREKTHGLVVENQELRQLGMDALVAEEBAEAKGN-EVRPVAGSAESAALRLRAPLQ 174
QY 180 QEGDEVPLDRI 190
Db 175 VQALSPLONI 185

```
RESULT 5
US-09-248-796A-14503
; Sequence 14503, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14503
; LENGTH: 1402
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-14503

Query Match      8.2%; Score 145.5; DB 2; Length 1402;
Best Local Similarity 23.5%; Pred. No. 0.00028;
Matches 73; Conservative 52; Mismatches 120; Indels 65; Gaps 14;

QY 47 QTKPEKKPAKRSKGQELVPKTNLPPKRAKTEDEKEQRIERVLNRRAAAQTSRER 106
Db 807 QKPEEHRKVEAQKAEAAK--LKEERRKAEERKQKEEKQKELLKKQKEEK 864
QY 107 KRLM-----EKLESEKIDMEQONQFL-----QRLAQMEANNRLSQVAQ--LSAE 152
Db 865 RQKELRKQREBEKEEAARLEERTKLMVNDDELAQIEVEKSKLSAAVANPLLNL 924
QY 153 VRGSRHSTPTSSPASVSPTLTP-----TLFKQEGDVPDLRIP----- 191
Db 925 YQSPGSAFTPTSTANL--PALSPLQASAKMLMQEQEQHQVLOEKLPQTSNIQLPNQ 983
QY 192 FPTPSV-----TDYSPTL-----KPSLSAESPOLTQHP--AVSYGGLGDS-----SALTIF 236
Db 984 QPHPSISSQFSSEYNPNASVFHNSSLLSNPSIMNSPRTTNNLNGNSPIVPNVTTNI 1043
QY 237 DIGASTKHEPHTDLTAPLDDDFRRLFN-----GSSLESDDSLLEDGFAPDVLDSGDL 290
Db 1044 SLGAT-----NTSNLSPSSKRLNSLSNSTQPPFGGNQFTQTNTASFGVNAVOQSGNF 1099
QY 291 SAPPFDSMVD 300
Db 1100 S--PFNAFSD 1107

RESULT 6
US-08-721-684C-2
; Sequence 2, Application US/08721684C
; Patent No. 5854016
; GENERAL INFORMATION:
; APPLICANT: Keegan, Kathleen S.
; TITLE OF INVENTION: No. 5854016el CREBA Isoform
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/721,684C

Query Match      8.0%; Score 141.5; DB 1; Length 521;
Best Local Similarity 24.0%; Pred. No. 0.00015;
Matches 82; Conservative 48; Mismatches 131; Indels 81; Gaps 16;

QY 8 SPVKMEDAFANSPPTTSPLEVPVLTVPADTSLRTKNVVAQTKPEEKPAKRSKGQELP 67
Db 234 SPVR---AMPRGPSALSTS-PLLT---APHKLOGSGPLVLTBEE-----KRTLVAEGYP 280
QY 68 VPKNLPPKRAKTEDEKEQRIERVLNRRAAAQTSRERKRLMEKLESEKIDMEQONQF 127
Db 281 IP-TKLPLTK---SEKALKKIRKIKNKISAQESRRKKKQKYMDSLEKKVESCSTENLE 335
QY 128 LLQRLAQMEANNRLSQVAQLSAEVRGSRHSTPTSSPASVSPTLTPT----- 176
Db 336 LRKKVEVLNTRTLQQLQKLTVMGK-----VSRCKLAGTGTGCLMVVLCFAV 389
QY 177 ----LFKQEGDVPDLRIPFPTSPVTDYSPTLKPSLSAESPOLTQHPAVSYGGLGDESA 232
Db 390 AFGSFGQYGV-----PYPSATKMA---LPS-----QHP-----LSEPYTA 421
QY 233 LTLFDLGLASI--KHBPHTDLTAPLDDDFRRLFNPGNDSSLESDDSLLEDGFAPDVLDSGDL 290
Db 422 SYVRSRNLIIYBEHAPLEESSPASGTGELGGWDRGSSLLRASSGL-----EALPEVDL 474
QY 291 SAPPFDSMVDFTPEPVTLEDLEQ---TNGLSDSASCKAASLQ 329
Db 475 PHFLISN---ETSLEKSVLLEQLQHLVSSKLEGNETILKVVELE 514

RESULT 7
US-09-005-970-2
; Sequence 2, Application US/09005970
; Patent No. 5959079
; GENERAL INFORMATION:
; APPLICANT: Keegan, Kathleen S.
; TITLE OF INVENTION: No. 5959079el CREBA Isoform
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/005,970
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/721,684
; FILING DATE:
```

```
/
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Williams Jr., Joseph A.
/ REGISTRATION NUMBER: 38,659
/ REFERENCE/DOCKET NUMBER: 27866/33487
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 312-474-6300
/ TELEFAX: 312-474-0448
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 521 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-005-970-2

Query Match      8.0%; Score 141.5; DB 1; Length 521;
Best Local Similarity 24.0%; Pred. No. 0.00015;
Matches 82; Conservative 48; Mismatches 131; Indels 81; Gaps 16;

QY 8 SPVKMEDAFANSPPTPSLEVPVLTVPSPADTSLRTKNVVAQTKPEKKPAKKKSWGOELP 67
Db 234 SPVR---AMPRGPSALSTS-PLLT---APHKLOGSGPLVLTTEE-----KKTLLVAEGYP 280
QY 68 VPKNLPPRKRAKTEDEKEQRRIERVLNRNRAAAQTSRRKRLEMEKLESEKIDMQNQF 127
Db 281 IP-TKLPLTK----SEKALKKIRKIKNKISAQESRRKKKEYMDSLEKKVBCSTENLE 335
QY 128 LLQRLAQMAEANNRLSQVAQLSAEVRGSRHSTPTSSPASVSPILTPT-----176
Db 336 LRKKVEVLNRTLLQQLQKLQTLVMGK-----VSRCKLAGTGTGCLMVMVVLCPAV 389
QY 177 ----LFKQEGDEVPLDRIPFPPTSPVTDYSPTLKPSSLAESPDLTQHPAVSVGLEGDESA 232
Db 390 AFGSFFQGYG-----PYPSTAKVA---LPS-----QHP-----LSEPYTA 421
QY 233 LTLFDLGASI--KHEPTHDLTAPLSDDDFRRLFNQDSSLESDDLLEGFADFVLDGDL 290
Db 422 SVVSRNLLIYEHAPLEBSSSPASTGELGGWDRGSSLLRASSGL-----EALPEVDL 474
QY 291 SAFFPDSMVDFTPEPTLEDLEQ---TNGLSDSASCKAASLQ 329
Db 475 PHFLISN--ETSLEKSVLLEQLQHLVSSKLEGNETLKVVELE 514

RESULT 8
US-09-407-715-2
/ Sequence 2, Application US/09407715
/ Patent No. 6248532
/ GENERAL INFORMATION:
/ APPLICANT: Keegan, Kathleen S.
/ TITLE OF INVENTION: No. 6248532e1 CREBa Isoform
/ NUMBER OF SEQUENCES: 10
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Marshall, O'Toole, Gerstein, Murray & Borun
/ STREET: 233 South Wacker Drive, 6300 Sears Tower
/ CITY: Chicago
/ STATE: Illinois
/ COUNTRY: USA
/ ZIP: 60606
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ APPLICATION NUMBER: US/09/407,715
/ FILING DATE: 28-Sep-1999
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/09/005,970
/ FILING DATE: 12-JAN-1998
/ APPLICATION NUMBER: US/08/721,684
/ FILING DATE: 27-SEP-1996
/ ATTORNEY/AGENT INFORMATION:
/
```

```
/
/ NAME: Williams Jr., Joseph A.
/ REGISTRATION NUMBER: 38,659
/ REFERENCE/DOCKET NUMBER: 27866/33487
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 312-474-6300
/ TELEFAX: 312-474-0448
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 521 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-407-715-2

Query Match      8.0%; Score 141.5; DB 2; Length 521;
Best Local Similarity 24.0%; Pred. No. 0.00015;
Matches 82; Conservative 48; Mismatches 131; Indels 81; Gaps 16;

QY 8 SPVKMEDAFANSPPTPSLEVPVLTVPSPADTSLRTKNVVAQTKPEKKPAKKKSWGOELP 67
Db 234 SPVR---AMPRGPSALSTS-PLLT---APHKLOGSGPLVLTTEE-----KKTLLVAEGYP 280
QY 68 VPKNLPPRKRAKTEDEKEQRRIERVLNRNRAAAQTSRRKRLEMEKLESEKIDMQNQF 127
Db 281 IP-TKLPLTK----SEKALKKIRKIKNKISAQESRRKKKEYMDSLEKKVBCSTENLE 335
QY 128 LLQRLAQMAEANNRLSQVAQLSAEVRGSRHSTPTSSPASVSPILTPT-----176
Db 336 LRKKVEVLNRTLLQQLQKLQTLVMGK-----VSRCKLAGTGTGCLMVMVVLCPAV 389
QY 177 ----LFKQEGDEVPLDRIPFPPTSPVTDYSPTLKPSSLAESPDLTQHPAVSVGLEGDESA 232
Db 390 AFGSFFQGYG-----PYPSTAKVA---LPS-----QHP-----LSEPYTA 421
QY 233 LTLFDLGASI--KHEPTHDLTAPLSDDDFRRLFNQDSSLESDDLLEGFADFVLDGDL 290
Db 422 SVVSRNLLIYEHAPLEBSSSPASTGELGGWDRGSSLLRASSGL-----EALPEVDL 474
QY 291 SAFFPDSMVDFTPEPTLEDLEQ---TNGLSDSASCKAASLQ 329
Db 475 PHFLISN--ETSLEKSVLLEQLQHLVSSKLEGNETLKVVELE 514

RESULT 9
US-09-350-841A-1601
/ Sequence 1601, Application US/09350841A
/ Patent No. 6750008
/ GENERAL INFORMATION:
/ APPLICANT: Jeffs, Peter;
/ TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION OF MEMBRANE
/ FUSION-ASSOCIATED EVENTS, INCLUDING HIV TRANSMISSION
/ FILE REFERENCE: 7872-066-999
/ CURRENT APPLICATION NUMBER: US/09/350,841A
/ CURRENT FILING DATE: 1999-07-09
/ NUMBER OF SEQ ID NOS: 1946
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 1601
/ LENGTH: 340
/ TYPE: PRT
/ ORGANISM: Human immunodeficiency virus type 1
US-09-350-841A-1601

Query Match      8.0%; Score 141; DB 2; Length 340;
Best Local Similarity 23.1%; Pred. No. 8.5e-05;
Matches 80; Conservative 40; Mismatches 127; Indels 100; Gaps 13;

QY 4 ADRF----SPVKMEDAFANSPPTPSLEVPVLTVPSPADTSLRTKNVVAQTKPEKKPAKKR 59
Db 35 ADSFSSMGSPVNAQDFCTDLAVSSANFIP--TVTAISTSPDLQWLQVPALVSSVAPSQTR 92
QY 60 KSWGQELPVKTNLPPR-----KKAQTE---DEKEQRRIERVLNRNRAA 100
/
```



```
Db 93 APHPFCVPAPSAGYBAGVVKWTGGAQSIGRRKVKVEQLSEEEKRRIRRE-RNKMAA 151
Qy 101 QTSREKRKLEMEKLESEKIDMEQOQFLQRLQAEAEENRLS-----143
Db 152 AKCRNRRLTDTLOAETDQLEDEKALQTEIANLLKEKEKLEFILAARHPACKIPDDLG 211
Qy 144 --QOVAQLAEVRGSRHSTSSPASVPTLTPFKQBGDRVLDRIPTTPSVTDYS 201
Db 212 PFEEMVASLDLFG---GLPEVATPS-----BEAFTPLNDPEPKPSV---E 254
Qy 202 PTLKPSL---AESPDLOHPAVSVGGLGDESAITLFDLGASIKHEPTHTDLPASL--- 255
Db 255 PVKISSMELKTEPPDPSPVPEVKSISSME-----LKTEPDDFLFPASSRP 300
Qy 256 -----DDFRRLFGDSLSLSDSLLEDGFAPDVLDGDLGA 292
Db 301 SGSETARSVPDMDLGSFYAGSSNRPSS-----DSLSPPTLA 339

RESULT 10
US-09-640-211A-1930
; Sequence 1930, Application US/09640211A
; Patent No. 6833446
; GENERAL INFORMATION:
; APPLICANT: Wood, Marion
; APPLICANT: Shenk, Michael A.
; APPLICANT: McGrath, Annette
; APPLICANT: Glenn, Matthew
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Modification of Gene Transcription
; FILE REFERENCE: 11000.1021C1U
; CURRENT APPLICATION NUMBER: US/09/640,211A
; CURRENT FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 2368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1930
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-640-211A-1930

Query Match 7.8%; Score 137.5; DB 2; Length 143;
Best Local Similarity 24.8%; Pred. No. 4.8e-05;
Matches 38; Conservative 25; Mismatches 45; Indels 45; Gaps 2;

Qy 18 NSPTTPSLEVPVLTSPADTSLRTKNVVAQTKPEKKPAKRKSMGQELPVKTNLPPRK 77
Db 5 SSVSSPDPTPSQSAAVRPTSTRDSSVME-----PPRK 38
Qy 78 RAKTEDEKQRIERVLRNRAAQTSSRKRLEMEKLESEKIDMEQOQFL-----128
Db 39 RARADLNARQREARAHNRNRIAAQNSRDKRKAQFTYMEQORVAQLEENQRLRAGMGLSQF 98
Qy 129 -----LORLAQMEENRLSQOVAQLSA 151
Db 99 TPADNDKFVSLERESQVARENRELKERIKSLES 131

RESULT 11
US-09-247-155-113
; Sequence 113, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bouqueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET-021A
; CURRENT APPLICATION NUMBER: US/09/247,155A
; CURRENT FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/081,563

Db 93 APHPFCVPAPSAGYBAGVVKWTGGAQSIGRRKVKVEQLSEEEKRRIRRE-RNKMAA 151
Qy 101 QTSREKRKLEMEKLESEKIDMEQOQFLQRLQAEAEENRLS-----143
Db 152 AKCRNRRLTDTLOAETDQLEDEKALQTEIANLLKEKEKLEFILAARHPACKIPDDLG 211
Qy 144 --QOVAQLAEVRGSRHSTSSPASVPTLTPFKQBGDRVLDRIPTTPSVTDYS 201
Db 212 PFEEMVASLDLFG---GLPEVATPS-----BEAFTPLNDPEPKPSV---E 254
Qy 202 PTLKPSL---AESPDLOHPAVSVGGLGDESAITLFDLGASIKHEPTHTDLPASL--- 255
Db 255 PVKISSMELKTEPPDPSPVPEVKSISSME-----LKTEPDDFLFPASSRP 300
Qy 256 -----DDFRRLFGDSLSLSDSLLEDGFAPDVLDGDLGA 292
Db 301 SGSETARSVPDMDLGSFYAGSSNRPSS-----DSLSPPTLA 339

RESULT 10
US-09-640-211A-1930
; Sequence 1930, Application US/09640211A
; Patent No. 6833446
; GENERAL INFORMATION:
; APPLICANT: Wood, Marion
; APPLICANT: Shenk, Michael A.
; APPLICANT: McGrath, Annette
; APPLICANT: Glenn, Matthew
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Modification of Gene Transcription
; FILE REFERENCE: 11000.1021C1U
; CURRENT APPLICATION NUMBER: US/09/640,211A
; CURRENT FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 2368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1930
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-640-211A-1930

Query Match 7.8%; Score 137.5; DB 2; Length 143;
Best Local Similarity 24.8%; Pred. No. 4.8e-05;
Matches 38; Conservative 25; Mismatches 45; Indels 45; Gaps 2;

Qy 18 NSPTTPSLEVPVLTSPADTSLRTKNVVAQTKPEKKPAKRKSMGQELPVKTNLPPRK 77
Db 5 SSVSSPDPTPSQSAAVRPTSTRDSSVME-----PPRK 38
Qy 78 RAKTEDEKQRIERVLRNRAAQTSSRKRLEMEKLESEKIDMEQOQFL-----128
Db 39 RARADLNARQREARAHNRNRIAAQNSRDKRKAQFTYMEQORVAQLEENQRLRAGMGLSQF 98
Qy 129 -----LORLAQMEENRLSQOVAQLSA 151
Db 99 TPADNDKFVSLERESQVARENRELKERIKSLES 131

RESULT 11
US-09-247-155-113
; Sequence 113, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bouqueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET-021A
; CURRENT APPLICATION NUMBER: US/09/247,155A
; CURRENT FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/081,563
```

```
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/099,273
; EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 113
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -310...-1
US-09-247-155-113

Query Match 7.7%; Score 135.5; DB 2; Length 395;
Best Local Similarity 25.2%; Pred. No. 0.00035;
Matches 62; Conservative 38; Mismatches 105; Indels 41; Gaps 10;

Qy 27 VPVLTVSPADTSLRTKNVVAQTKPEKKPAKRKSMGQELPVKTNLPPRKRAKTEDEKE 86
Db 176 VPCTLLPCQTLFLT-----DEEK-----RLGQE-----GVSLPSHLPLTKAERV 217
Qy 87 QRRIERVLRNRAAQTSSRKRLEMEKLESEKIDMEQOQFLQRLQAEAEENRLSQOV 146
Db 218 LKKVRKRIRNKQSAQDSRRRKEYIDGLSRVAACSAQNQELQKVQELERHNISLVAQL 277
Qy 147 AQLSAEVRGSRHSTSSPASVPTLTPFKQBGDEVPDLRIPE---PTPSVTDYSP- 202
Db 278 RLQQLTI-----AQTSNKAQTSICVILLFSLALIILP-SFSPQSPRPEAGSEYQPH 330
Qy 203 TLKPSSLAESPDLOHPAVSV-----GGLEGDESALTIFD-LGASIKHEPTHTDLTA 252
Db 331 GVTSRNLTHTKQDVNTENLETQVVSRLRBPFGAKDANGSTRITLLEKMG--KRPSPGRIRS 388
Qy 253 PLSDD 258
Db 389 VLHADE 394

RESULT 12
US-09-513-999C-14
; Sequence 14, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 14
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -310...-1
; OTHER INFORMATION: score 4.4
; OTHER INFORMATION: seq VLIILFSLALIIL/PS
US-09-513-999C-14

Query Match 7.7%; Score 135.5; DB 2; Length 395;
Best Local Similarity 25.2%; Pred. No. 0.00035;
Matches 62; Conservative 38; Mismatches 105; Indels 41; Gaps 10;
```


Search completed: November 23, 2005, 03:27:33
Job time : 35.1763 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:05:10 ; Search time 149.973 Seconds
(without alignments)
1641.819 Million cell updates/sec

Title: US-10-663-450-4

Perfect score: 1765

Sequence: 1 MKSADRFSPVKMEDAFANSP.....PSHGASTSRCDGQGIAGSA 349

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1755.5	99.4	350	2	Q8TFU8_EMENI
2	1696	96.0	347	2	Q5AQN3_EMENI
3	1293.5	73.2	342	2	Q6W8X2_ASPNG
4	856	48.5	433	2	Q4WEY8_ASPFU
5	434	24.6	451	2	Q8TFP3_TRIRE
6	423	24.0	430	2	Q7SHF0_NEUCR
7	412.5	23.4	429	2	Q4HT55_GIBZE
8	346	19.6	556	2	Q5IKW8_MAGGR
9	261.5	14.8	299	2	Q6CEV1_YARLI
10	217	12.3	273	2	Q6CKQ1_KJULA
11	209.5	11.9	260	2	Q6BQC2_DEBHA
12	208	11.8	357	2	Q5AA52_CANAL
13	206.5	11.7	230	1	HACL_YEAST
14	177.5	10.1	396	2	Q7ZYC2_XENLA
15	169	9.6	228	2	Q75BQ5_ASHGO
16	169	9.6	321	2	Q8GRY7_LOTJA
17	168.5	9.5	350	2	Q90ZRT_XENLA
18	167.5	9.5	336	2	Q5VR11_ORYSA
19	166.5	9.4	437	2	Q569T3_XENLA
20	166	9.4	458	2	Q4P9B7_USTMA
21	166	9.4	639	2	Q6CET1_YARLI
22	165.5	9.4	322	2	Q39896_SOYBN
23	164.5	9.3	176	2	Q69XK6_ORYSA
24	164.5	9.3	634	2	Q59W43_CANAL
25	164.5	9.3	634	2	Q59W83_CANAL
26	163.5	9.3	168	1	HIS_ARATH
27	163.5	9.3	470	2	Q5FVM5_RAT
28	163	9.2	252	2	Q6NX18_XENTR
29	163	9.2	326	2	Q39895_SOYBN
30	162	9.2	322	2	O04234_VICFA
31	161.5	9.1	329	2	Q6FLY3_CANGA

32	161.5	9.1	479	2	Q91XE9_MOUSE
33	161	9.1	585	2	Q4P8Q4_USTMA
34	160.5	9.1	404	1	CREB3_MOUSE
35	160	9.1	261	1	XBPI_HUMAN
36	159.5	9.0	807	2	Q51LL3_MAGGR
37	159	9.0	721	2	O22208_ARATH
38	159	9.0	721	2	O8L7E7_ARATH
39	158.5	9.0	479	2	Q8BWS0_MOUSE
40	157	8.9	616	2	Q524G5_MAGGR
41	156	8.8	277	2	Q90513_FUGRU
42	156	8.8	378	2	Q800A6_FUGRU
43	154.5	8.7	686	2	O61D33_CAEBR
44	154.5	8.7	690	2	O44743_CAEBL
45	153.5	8.7	327	1	FOSL2_RAT

ALIGNMENTS

RESULT 1

Q8TFU8_EMENI PRELIMINARY; PRT; 350 AA.
AC Q8TFU8;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Transcription factor.
GN Name=hacA;
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Saloheimo M.L.A., Valkonen M., Penttilae M.E.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ413273; CAC88375.1; -; Genomic_DNA.
DR HSSP; P05412; IJNM.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; BZIP_2.
DR InterPro; IPR004827; TF_BZIP.
DR Pfam; PF07716; BZIP_2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS00217; BZIP; 1.
DR PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 350 AA; 38192 MW; 4F78340096064F44 CRC64;

Query Match 99.4%; Score 1755.5; DB 2; Length 350;
Best Local Similarity 99.7%; Pred. No. 7.5e-83;
Matches 349; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY	1	MKSADRFSPVKMEDAFANSP-PTTPSLEVPVLTSPADTSLRTKNVVAQTKPEKKPAKRR	59
DB	1	MKSADRFSPVKMEDAFANSPPTTPSLEVPVLTSPADTSLRTKNVVAQTKPEKKPAKRR	60
QY	60	KSWGQELVPKTNLPKRAKTEDEKQRRIRVLRNRAAQTSTRERKLEMEKLESEKI	119
DB	61	KSWGQELVPKTNLPKRAKTEDEKQRRIRVLRNRAAQTSTRERKLEMEKLESEKI	120
QY	120	DMEQNQQLLQRLAQMEANNRLSQVLAQSLAEVGRSHSTPTSSPASVPTLTPTLFK	179
DB	121	DMEQNQQLLQRLAQMEANNRLSQVLAQSLAEVGRSHSTPTSSPASVPTLTPTLFK	180
QY	180	QEGDEVPIDRIPFPPTSVTDYSPTLKPSLAESPDLTOHPAVSVGLEGDEGSAITFDLG	239
DB	181	QEGDEVPIDRIPFPPTSVTDYSPTLKPSLAESPDLTOHPAVSVGLEGDEGSAITFDLG	240
QY	240	ASIKHEPHTDLTAPLSDDDFRFLFNGDSLSLLEDGFADFVLDSDGLSAFPFDSMW	299
DB	241	ASIKHEPHTDLTAPLSDDDFRFLFNGDSLSLLEDGFADFVLDSDGLSAFPFDSMW	300

QY 300 DFDTEPVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDGGIAAGSA 349
 |||||
 Db 301 DFDTEPVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDGGIAAGSA 350
 |||||

RESULT 2

Q5AQN3 EMENI PRELIMINARY; PRT; 347 AA.
 ID Q5AQN3 ASPNG
 AC Q6W8X2;
 DT 10-MAY-2005 (TremBLrel. 30, Created)
 DT 10-MAY-2005 (TremBLrel. 30, Last sequence update)
 DE Hypothetical protein.
 GN ORFName=AN9397.2;
 OS Aspergillus nidulans FGSC A4.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; Emericella.
 OX NCBI_TaxID=227321;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=FGSC A4;
 RA Birren B., Nuebaum C., Abouelleil A., Allen N., Anderson S.,
 RA Arachchi H.M., Barna N., Baetien V., Bloom T., Boguslavskiy L.,
 RA Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
 RA Choepel Y., Collymore A., Cook A., Cooke P., Corum B., DeArellano K.,
 RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
 RA Erickson J., Faro S., Ferreira P., FitzGerald M., Gage D., Galagan J.,
 RA Gardyna S., Gnerre S., Graham L., Grand-Pierre N., Hafez N.,
 RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,
 RA Jaffe D., Johnson R., Jones C., Kamal M., Katat A., Karatas A.,
 RA Kella C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
 RA Ma L.-J., Mabbitt R., MacLean C., Macdonald P., Major J., Manning J.,
 RA Matthews C., Maucelli E., McCarthy M., Meldrim J., Meneus L.,
 RA Mihova T., Mienga N., Murphy T., Naylor J., Nguyen C., Nicol R.,
 RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neil D.,
 RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,
 RA Rachupka A., Ramasamy U., Raymond C., Retta R., Riese C., Rogov P.,
 RA Roman J., Schauer S., Schupack R., Seaman S., Severy P., Smirnov S.,
 RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,
 RA Talamas J., Tesfaye S., Theodore J., Topham K., Travers M.,
 RA Vassiliev H., Venkataranan V.S., Viel R., Vo A., Wang S., Wilson B.,
 RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
 RA Lander E.;
 RT "Genome Sequence of Aspergillus nidulans";
 RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AACD01000172; EAA66464.1; -; Genomic_DNA.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 KW DNA-binding; Hypothetical protein; Nuclear protein.
 SQ SEQUENCE 347 AA; 37841 MW; A7CNAFEE761B3E1 CRC64;

Query Match 96.0%; Score 1696; DB 2; Length 347;
 Best Local Similarity 97.4%; Pred. No. 8.6e-80;
 Matches 341; Conservative 1; Mismatches 4; Indels 4; Gaps 2;

QY 1 MKSADRFSPVKMEDAFANS-PTTPSLSEVPLVTVSPADTSLRTKNVVAQTKPEKPAK 59
 |||||
 Db 1 MKSADRFSPVKMEDAFANSPTTPSLSEVPLVTVSPADTSLQTKNVVAQTKPEKPAK 60
 |||||
 QY 60 KSWGELPVKTNLPKRAKTEDEKEQRRIERVLNRAAAQTSRERKLEMEKLESEKI 119
 |||||
 Db 61 KSWGELPVKTNLPKRAKTEDEKEQRRIERVLNRAAAQTSRERKLEMEKLESEKI 120
 |||||
 QY 120 DMEQNNQFLQRLAQMEANRLSQVAQLSAEVRGSRHSTPTSSPASVSTLTPTLFK 179
 |||||
 Db 121 DMEQNNQFLQRLAQMEANRLSQVAQLSAEVRGSRHSTPTSSPASVSTLTPTLFK 180
 |||||
 QY 180 QSGDEVPLDRIIPFPPTSVTDYPTLPKPSLSAEPDLTQHPAVSVGGLGDESALTFLD 239
 |||||
 Db 181 QSGDEVPLDRIIPFPPTSVTDYPTLPKPSLSAEPDLTQHPAVSVGGLGDESALTFLD 237
 |||||

QY 240 ASIKHEPTHDLTAPLSDDDFRRLFNGSDSSLESDDSLLEDGFADVLDSGLSAFFPDSMV 299
 |||||
 Db 238 ASIKHEPTHDLTAPLSDDDFRRLFNGSDSSLESDDSLLEDGFADVLDSGLSAFFPDSMV 297
 |||||
 QY 300 DFDTEPVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDGGIAAGSA 349
 |||||
 Db 298 DFDTEPVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDGGIAAGSA 347
 |||||

RESULT 3

Q6W8X2 ASPNG PRELIMINARY; PRT; 342 AA.
 ID Q6W8X2;
 AC Q6W8X2;
 DT 05-JUL-2004 (TremBLrel. 27, Created)
 DT 05-JUL-2004 (TremBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TremBLrel. 27, Last annotation update)
 DE Transcription factor HACA.
 GN Name=hacA;
 OS Aspergillus niger.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=5061;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Mulder H.J.;
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY303684; AAO73495.1; -; Genomic_DNA.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR011700; bZIP 2.
 DR InterPro; IPR004827; TF bZIP.
 DR Pfam; PF07116; bZIP_2; 1.
 DR SMART; SM00338; BRLZ; 1.
 DR PROSITE; PS0217; BZIP; 1.
 DR PROSITE; PS0036; BZIP_BASIC; UNKNOWN_1.
 KW DNA-binding; Nuclear protein.
 SQ SEQUENCE 342 AA; 37148 MW; 623B3941A55C9C05 CRC64;

Query Match 73.2%; Score 1293.5; DB 2; Length 342;
 Best Local Similarity 76.2%; Pred. No. 4.6e-59;
 Matches 262; Conservative 29; Mismatches 46; Indels 7; Gaps 3;

QY 5 DRFSFVKMEDAFANSPTTPSLSEVPLVTVSPADTSLRTKNVVA-QTKPEKKPAKRRKSMG 63
 |||||
 Db 4 EAFSPV---DSLGSFTP---ELPLTVSPADTSLDSSVQAGETKAEKKPKVKKKSMG 57
 |||||
 QY 64 QELPVKTNLPKRAKTEDEKEQRRIERVLNRAAAQTSRERKLEMEKLESEKIDMEQ 123
 |||||
 Db 58 QELPVKTNLPKRAKTEDEKEQRRIERVLNRAAAQTSRERKLEMEKLESEKIDMEQ 117
 |||||
 QY 124 QNOFLQRLAQMEANRLSQVAQLSAEVRGSRHSTPTSSPASVSTLTPTLPKQEGD 183
 |||||
 Db 118 QNOFLQRLAQMEANRLNQVAQLSAEVRGSRGTRPKGSPVASPTLTPTLPKQERD 177
 |||||
 QY 184 EYPLDRIIPFPPTSVTDYPTLPKPSLSAEPDLTQHPAVSVGGLGDESALTFLDGASIK 243
 |||||
 Db 178 EYPLERIPFPPTSIIDYPTLPSTLAESSDVTQHPAVSVAGLEGGSALSFDVGSNPE 237
 |||||
 QY 244 HEPTHDLTAPLSDDDFRRLFNGSDSSLESDDSLLEDGFADVLDSGLSAFFPDSMVDFDT 303
 |||||
 Db 238 PHAADLAAPLSDDDFRRLFNVDSVPVGSDDSVLEDDGFADVLDSGLSAFFPDSMVDFDP 297
 |||||
 QY 304 EPVTLLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDGGIAAG 347
 |||||
 Db 298 ESVGEGIEPHGLPDETSTROTSSVQPSLGASTRCDGGIAAG 341
 |||||

RESULT 4

Q4WEY8 ASPFU PRELIMINARY; PRT; 433 AA.
 ID Q4WEY8 ASPFU
 AC Q4WEY8;
 DT 13-SEP-2005 (TremBLrel. 31, Created)

ID	QTFE3_TIRE PRELIMINARY;	PRT;	451 AA.
AC	QTFE3;		
DT	01-JUN-2002 (TrEMBLrel. 21, Created)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)		
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)		
DE	Transcription factor.		
GN	Nameshaci;		
OS	Trichoderma reesei (Hypocrea jecorina).		
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;		
OC	Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.		
OX	NCBI_TaxID=51453;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RA	Salonheimo M.L.A.; Valkonen M.; Penttilae M.E.;		
RL	Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AJ141372; CAC88374.1; -; Genomic_DNA.		
DR	GO; GO:0005634; C:nucleus; IEA.		
DR	GO; GO:0003677; F:DNA binding; IEA.		
DR	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.		
DR	InterPro; IPR011700; bZIP 2.		
DR	InterPro; IPR004827; TF bZIP.		
DR	Pfam; PF07116; bZIP 2; 1.		
DR	SMART; SM00338; BRLZ; 1.		
DR	PROSITE; PS50217; bZIP; 1.		
KW	DNA-binding; Nuclear protein.		
SQ	SEQUENCE 451 AA; 49277 MW; 600F10B471EA3AD3 CRC64;		
Query Match 24.6%; Score 434; DB 2; Length 451;			
Best Local Similarity 32.4%; Pred. No. 1.1e-14;			
Matches 149; Conservative 50; Mismatches 109; Indels 152; Gaps 21;			
Qy	5	DRFSPVKMDAFANSTTPSLEVPVLTVSPADTSLRTKNVVAQ-----	47
Db	26	DNFTSL-----PADS--TPS-----TLNPRD--MMTPDSVADISRLSVIPESQDAEDD	70
Qy	48	-----TKP--BEKPAKRRKSWGELPVPKTNLPKRKAKTEDEKEQRRIRVLENRA	98
Db	71	ESHSTSATAPSTSEKPKVKKRKSQGVLPKPKTNLPKRKAKTEDEKEQRRIRVLENRR	130
Qy	99	AAOTSRERKRLMEKLESEK-----IDMEQOQFLQLORLAQEAENNRLS-----	143
Db	131	AAQSSRRKRLVEALEKRNKELETLLINQVKNLILVELNRFRRSSGVVTRSSSPLDS	190
Qy	144	-QQVAQLSAEVRGSRH-----STPTSS-----SPASVSPTLTPTLFKQ----	180
Db	191	LQDSITLSQQLFSGRDGQWMSNPESQLMDQIMRSAANPTVNPASLSPSPISDKKEFTK	250
Qy	181	EGDEVPLRIPTPTPSVTYDYSPTLKSSLA-----SPDLTQHPAVSVGLEGDESA	232
Db	251	EEDEEQADE---DEENEQWTHETKEAAAAKEKNSKOSRVSTDTORPAVSIGG----DAA	303
Qy	233	LTLLF--DLGAS-IKHPTHTDLTAPLSDDDFRRLFNGDSLSLSDSSILEGFAFDVLDSGD	289
Db	304	VPVFSODAGANCIGLGFVHQDGPFF---IGHFGLSALDADRYLLE-----SQ	350
Qy	290	LSAFPFDMSVDFD-----TEEVTLLEDTQNLGLSDSASCKAASL-----	328
Db	351	LLASPNASTVDDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAAD	410
Qy	329	-----QFSHGASTRCDGGQIAAG	347
Db	411	RELDLEIHPENQIPSRHSITQPPQSGASSHGCDGGIAVG	450
RESULT 6			
Q7SHF0_NEUCR			
ID	Q7SHF0_NEUCR PRELIMINARY;	PRT;	430 AA.
AC	Q7SHF0;		
DT	01-MAR-2004 (TrEMBLrel. 26, Created)		
DT	01-MAR-2004 (TrEMBLrel. 26, Last sequence update)		
DE	Predicted protein.		
GN	Name=NCU01856.1;		

```
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehm B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Sellitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamysseil M., Mauceli E., Bielke C., Rudd S., Friseman D.,
RA Krystofova S., Ramussen C., Metzenberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Omani S.A.,
RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B.,
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL Nature 0:0-0(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABX01000004; EAA36251.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; F:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; bZIP 2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07716; bZIP 2; I.
DR PROSITE; PS50217; bZIP; 1.
SQ SEQUENCE 430 AA; 46599 MW; 0EB0657CFA6160DA CRC64;

Query Match      24.0%; Score 423; DB 2; Length 430;
Best Local Similarity 33.3%; Pred. No. 4e-14;
Matches 139; Conservative 35; Mismatches 137; Indels 106; Gaps 13;

QY 3 SADRFSPVKMED-----AFNSPTTSPLEVPVITVSPADTSLRTKNVVAQTKPEEKKP 55
DB 46 SLDVMSPVSLADDDVLTALATLTVPQIPPRSTPASTPAP-----ETEEKP 90
QY 56 AKKRSWGQELVPKTNLPKRKTEDEKQRRVLRLNRAAAQTSRERKLEWEKLE 115
DB 91 VKKRSWGQVLPEPKTNLPKRKTEDEKQRRVLRLNRAAAQSSRERKRLVEGLE 150
QY 116 SEKIDME-----QONQFLQRLAQMEANNRLSQOVAQLSAEVRG----- 155
DB 151 RNNKELETLMQAQINOTLLQALRE-----NGVAPTIAATRPASFDGLNPTPTVPSQELF 205
QY 156 -----SRH-----STPTSSSPASVSTLTPTTLFKQEGDEVLPDRIPFPPTS 196
DB 206 SSQDGHNLKHDSSLEQLPPTIKTBETVNPASLSPVLNPLPEMEEDGEKEQSTAQPVAD 265
QY 197 VTDYSPTLKPSLAESPDLTQHPAV-----SVGLEGGDSALTPLDLGASIKHEPTH 248
DB 266 AT-----STATVDTSPDQHPAVVFPVAVSAPVAG-----SSISAAPLGLG---NEMD 311
QY 249 DLTAPLSDDDFRLFNQD--SSLESDSLLEDGFAFDVLDSDGSLAAPP---FDSMVDFDTE 304
DB 312 DLVLKTFDADRYIYETEFESFSPSNFDEYSWAGDDTETLHPANPENFFDEFINPDVS 371
QY 305 PVTLEDLEQTNGLSASCAASL-----QPSHGASTSRCDGGOGIAAG 347
DB 372 NAANSNEQPOQQSGFCATTNVAHVNLTYPNASEDPYQOPHTGASINGCDGGIAGV 428

RESULT 7
Q4HTT5_GIBZE
ID Q4HTT5_GIBZE PRELIMINARY; PRT; 429 AA.
AC Q4HTT5;
```



```

Db 344 IGDNNVGVDPAPFNLDNDFDISLWLNDDSAISAESNATSDFAAAIOGLEPKIYEPENQ-- 401
Qy 316 GLSDSASCAASLQPHGASTGRCDQGIAG 347
Db 402 -----VSSNPICQPHPGASTQCDVGGAIVG 428

RESULT 8
Q51KW8_MAGGR PRELIMINARY; PRT; 556 AA.
AC Q51KW8;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=MG09010.4;
OS Magnaporthe grisea 70-15.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes; Incertae sedis; Magnaportheaceae; Magnaporthe.
OX NCBI_TaxID=242507;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Birten B., Nuebaum C., Abebe A., Abouelleil A., Adekova E.,
RA Ait-zahra M., Allen N., Allen T., An P., Anderson M., Anderson S.,
RA Arachchi H., Ambruster J., Bachantsang P., Baldwin J., Barry A.,
RA Bayui T., Blishsteyn B., Bloom T., Blye J., Boguslavskiy L.,
RA Borowsky M., Boukhgalter B., Brunache A., Butler J., Calixte N.,
RA Calvo S., Camarata J., Campo K., Chang J., Cheshatsang Y., Citroen M.,
RA Collimore A., Considine T., Cook A., Cooke P., Corum B., Cuomo C.,
RA David R., Dawoe T., Degray S., Dodge S., Dooley K., Dorje P.,
RA Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T., Engels R.,
RA Fitzgeral J., Farina A., Faro S., Ferreira P., Fischer H.,
RA Fitzgerald M., Foley K., Gage D., Galagan J., Gearin G.,
RA Gierke A., Goyette A., Graham J., Grandbois E., Gyaltsen K., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,
RA Honan T., Horn A., Houde N., Hughes L., Hulme W., Hueb E., Iliev I.,
RA Jaffe D., Jones C., Kamal M., Kamat A., Kamysseilis M., Karlsson E.,
RA Kells C., Kieu A., Kisher P., Kodira C., Kulbokas E., Labutti K.,
RA Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,
RA Lindblad-toh K., Liu X., Lokyitsang T., Lokyitsang Y., Lucien O.,
RA Lui A., Ma L.J., Mabbitt R., Macdonald J., Maclean C., Major J.,
RA Manning J., Marabella R., Maru K., Matthews C., Mauceli E.,
RA McCarthy M., McDonough S., Mcghee T., Meldrim J., Meneus L.,
RA Mesirov J., Mihalev A., Mihova T., Mikkelson T., Mienga V., Moru K.,
RA Mozes J., Mulrain L., Munson G., Naylor J., Neves C., Nguyen C.,
RA Nguyen N., Nguyen T., Nicol R., Nielsen C., Nizzari M., Norbu C.,
RA Norbu N., O'donnell P., Okoawo O., O'leary S., Omotosho B.,
RA O'Neill K., Oman S., Parker S., Perrin D., Phunkhang P., Pigani B.,
RA Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond C.,
RA Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,
RA Rutman M., Schupbach R., Seaman C., Settipalli S., Sharpe T.,
RA Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,
RA Spencer B., Stalker J., Stange-thomann N., Stavropoulos S.,
RA Stenson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,
RA Tensing P., Tesfaye S., Theodore J., Thoultsang Y., Topham K.,
RA Towey S., Tsamila T., Tesomo N., Vallee D., Vassiliev H.,
RA Venkataraman V., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,
RA Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,
RA Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembeck L.,
RA Zimmer A., Zody M., Lander E.;
RT "The genome sequence of Magnaporthe grisea."
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Dean R., Mitchell T., Brown D., Pan H., Thon M.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Zhu H., Blackmon B.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.

```

```

CC -I- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AACU01001606; EAA47880.1; -; Genomic_DNA.
DR InterPro; IPR011700; bZIP 2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF0716; bZIP_2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS0217; bZIP; 1.
KW DNA-binding; Hypothetical protein; Nuclear protein.
SQ SEQUENCE 556 AA; 60798 MW; CFFC4C8D2DC4963C CRC64;

Query Match 19.6%; Score 346; DB 2; Length 556;
Best Local Similarity 36.3%; Pred.No. 4.9e-10;
Matches 103; Conservative 32; Mismatches 55; Indels 94; Gaps 12;

Qy 10 VKME-----DAFANSP--TTPSL--EVPVLTVSPAD-----TSRLTKN---- 43
Db 10 LKVEASPAESFLSTFGDMYPSLFGDAKFSNLPDLASPPGSDPVELDALKTESAALR 69
Qy 44 -VVAQTKPEE-----KKPAKKKSGQELVPKTNLPKPKRAKTEDEKQRRIVRL 94
Db 70 AVESTPTPEETSSPAAASEKTKKKGKSGQVLPKTNLPKPKRAKTEDEKQRRIVRL 129
Qy 95 RNRAAAQTSRERKLEMEKLESEKIDMEQNOFL--QLAQMEANRNRLSQOVAQLSAE 152
Db 130 RNRAAQSRRERKQVEALEQNALEQQ---LLHFQKLTQTMYOELQLRR----- 178
Qy 153 VRGSHSPPTSSPASVPTTLTFLFKQGEVPLDRIPFPFSPV----- 197
Db 179 ----REAGVTSSEKPDGLTLPFLRSQ-----DATASSATSVSATLADPANSLEDIF 228
Qy 198 --TVSPTLKPSLSAESP-----DLTQHPA 220
Db 229 RSTNVTTPVNPASIRSSPAPEQQHESVASGEAKTSADLTQHPA 272

RESULT 9
Q6CEV1_YARLI PRELIMINARY; PRT; 299 AA.
AC Q6CEV1;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Yarrowia lipolytica chromosome B of strain CLIB99 of Yarrowia
DE lipolytica.
GN OrderedLocusNames=YALIOB12716g;
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=CLIB 122 / E 150;
RX PubMed=15229592; DOI=10.1038/nature02579;
RA Dujon B., Sherman D., Fischer G., Dutrains P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Baray S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
RA Bolerame A., Boyer J., Catolico L., Confanieri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantreay F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.-M., Nikolaki M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,
RA Swennen D., Tekalia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RT "Genome evolution in yeasts."
RL Nature 430:35-44(2004).
DR EMBL; CR382128; CAG83062.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.

```


DR InterPro: IPR004827; TF bZIP.
 DR Pfam: PF07116; bZIP 2; 1.
 DR SMART: SM00338; BRU2; 1.
 DR PROSITE: PS00217; bZIP; 1.
 DR PROSITE: PS00036; bZIP BASIC; UNKNOWN 1.
 KW Complete proteome; DNA-binding; Nuclear protein.
 SQ SEQUENCE 260 AA; 29442 MW; 49B1819E393BFF6 CRC64;

Query Match 11.9%; Score 209.5; DB 2; Length 260;
 Best Local Similarity 28.8%; Pred. No. 0.0023;
 Matches 66; Conservative 41; Mismatches 61; Indels 61; Gaps 8;

QY 70 KTNLPKRKAKTEDEKEQRIERVLNRRAAAQTSRERKRLMEKLESEKIDME-----116
 DB 14 KSLPPKRKAKTEDEKEQRIERVLNRRAAAQTSRERKRLMEKLESEKIDME-----116
 QY 117 -----EKIDMEQGNQF-----LLQRLAQMEANNRLSQVLAQLSA 151
 DB 74 SNFDGVCQMSAKLAKIDMQEMDDLDLQDKLHTNLNGLLANEGDNDLDELTTPAE 133
 QY 152 EVRGRSHSTPTSSSPASVSFTLPTLFQEGDEGVPLDRIPF-----PTP-----SVTD--199
 DB 134 QPDTKRRLSSANSCSSLLATDLTESVEQVSSDKVNQIPHIKMEPNSDSNELLSTINEN 193
 QY 200 -YSPTLKPSL---AESP-DLT-----QHPAVSVCGLGDESAITLFD 237
 DB 194 VYNYLSPISINSPIIDLTLLKKSSENLPPLSLANNGDIDSLKYVD 242

RESULT 12
 Q5AA52 CANAL PRELIMINARY; PRT; 357 AA.
 AC Q5AA52;
 DT 10-MAY-2005 (TREMBLrel. 30, Created)
 DT 10-MAY-2005 (TREMBLrel. 30, Last sequence update)
 DE Hypothetical protein HAC1.
 GN Name=HAC1; ORFNames=CaO19.2432, CaO19.9968;
 OS Candida albicans SC5314.
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=237561;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=SC5314;
 RX PubMed=15123810; DOI=10.1073/pnas.0401648101;
 RA Jones T., Federspiel N.A., Chibana H., Dungan J., Kalman S.,
 RA Magee B.B., Newport G., Thorstenson Y.R., Agabian N., Magee P.T.,
 RA Davis R.W., Scherer S.;
 RT "The diploid genome sequence of Candida albicans."
 RL Proc. Natl. Acad. Sci. U.S.A. 101:7329-7334(2004).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=SC5314;
 RA Dungan J., Kuo A., Newport G., Lan C.-Y., Iijima C., Adegbola O.,
 RA Roberts J., Persson K., Donnelly S., Favoreto S., Tzung K.-W.,
 RA Jones T., Scherer S., Agabian N.;
 RT "Annotation of the Genome of Candida albicans."
 RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AACQ01000041; EAK9529.1; -; Genomic DNA.
 DR EMBL; AACQ01000040; EAK95617.1; -; Genomic DNA.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription; IEA.
 DR GO; GO:0006355; P:regulation of protein; Nuclear protein.
 KW DNA-binding; Hypothetical protein; Nuclear protein.
 SQ SEQUENCE 357 AA; 40127 MW; 8E1C633A2B4ECD4E CRC64;

Query Match 11.8%; Score 208; DB 2; Length 357;
 Best Local Similarity 23.0%; Pred. No. 0.0039;
 Matches 90; Conservative 61; Mismatches 141; Indels 100; Gaps 14;

QY 12 MEDAFANSTPTSPLEVPVLTSPATSLRTKVVVAQTPEEKPAKRSKQGLPVP--69
 DB 12 MEDAFANSTPTSPLEVPVLTSPATSLRTKVVVAQTPEEKPAKRSKQGLPVP--69

Db 1 MELTVNTNTTNSID-----DLSVATPTSLMTSTTTSPMSMTSTSSHSNTLIDPA 51
 QY 70 --KTNLPKRKAKTEDEKEQRIERVLNRRAAAQTSRERKRLMEKLESEKIDME-----122
 DB 52 TFKSTLPPKRAKTEDEKEQRIERVLNRRAAAQTSRERKRLMEKLESEKIDME-----111
 QY 123 -----QQNQFLQRLAQMEANNRLSQVLAQLSAEVRGRSHSTPTSSSPASV 169
 DB 112 LNNYNQAFELTKQNQELL--LSKLEVLDD--VSDLKEQTHSNMSTRRSHNKKSNDEDI 168
 QY 170 SPTLPTLPKQSGDEVPLDRIPPTSPVTDYSPTLKPSL-----AESPDLTQH 218
 DB 169 -----EEDDEHQEGHVEKQEIKEEPPVSKRKLNTYKTKTKSNKNTWST 219
 QY 219 PAVSVCGLGDE-----ESALTFLDGLGASIKHE-----PTHDLTAPLSD 256
 DB 220 PPSVSSLPDVTNFGTNTWSSPIQIKFENIDNIFIKKFSQSPLOPTHQ---PEQQ 276
 QY 257 DDFRLFNCGDSSLSLEDGFADVLDSGLSAPFPDSMVDPTPTVLEDTQNG 316
 DB 277 EQQHEIY-----LKSES---KDAF-----WNYPSPLSFHDSPLQI-DIDTSS 315
 QY 317 LSDSASCKAASLQPSHGASTSRCDGGIAAGS 348
 DB 316 SSTSPSSSSSLISSGPTNHSIADLAASLAS 347

RESULT 13
 HAC1 YEAST
 ID HAC1 YEAST STANDARD; PRT; 230 AA.
 AC P41546; P87040;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 13-SEP-2005 (Rel. 48, Last annotation update)
 DE HAC1 protein.
 GN Name=HAC1; Synonyms=ERN4, IRE2; OrderedLocustNames=YFL031W;
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetales; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=S288c;
 RX MEDLINE=95116316; PubMed=7816617;
 RA Nojima H., Leem S.-H., Araki H., Sakai A., Nakashima N., Kanaoka Y.,
 RA Ono Y.;
 RT "Hac1: a novel yeast bZIP protein binding to the CRE motif is a
 RT multicopy suppressor for cdc10 mutant of Schizosaccharomyces pombe."
 RL Nucleic Acids Res. 22:5279-5288(1994).
 RN [2]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=S288c / AB972;
 RX MEDLINE=95400292; PubMed=7670463;
 RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
 RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
 RA Yanazaki M.-A., Tashiro H., Eki T.;
 RT "Analysis of the nucleotide sequence of chromosome VI from
 RL Saccharomyces cerevisiae."
 RN Nat. Genet. 10:261-268(1995).
 RN [3]
 RP SEQUENCE REVISION TO 183-230.
 RA Murakami Y.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=AB320;
 RX MEDLINE=97222447; PubMed=9077435;
 RA Mori K., Kawahara T., Yoshida H., Yanagi H., Yura T.;
 RT "Signalling from endoplasmic reticulum to nucleus: transcription
 RT factor with a basic-leucine zipper motif is required for the unfolded
 RT protein-response pathway."
 RL Genes Cells 1:803-817(1996).
 RN [5]

RP CHARACTERIZATION.
RX MEDLINE=97086687; PubMed=8933376; DOI=10.1093/nar/24.21.4222;
RA Nikawa J.-I., Akiyoshi M., Hirata S., Fukuda T.;
RT "Saccharomyces cerevisiae IRE2/HAC1 is involved in IRE1-mediated KAR2
RT expression";
RL Nucleic Acids Res. 24:4222-4226(1996).
RN [6]
RN CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RP MEDLINE=97053779; PubMed=8898193; DOI=10.1016/S0092-8674(00)81360-4;
RA Cox J.S., Walter P.;
RT "A novel mechanism for regulating activity of a transcription factor
RT that controls the unfolded protein response";
RL Cell 87:391-404(1996).
RN -1- FUNCTION: Seems to be involved in the unfolded protein response
RN (UPR) pathway. Binds to the UPR element (UPRE) in the promoter of
RN UPR-regulated genes such as KAR2, PDI1, EUG1 and FKBP2; activates
RN the transcription of these genes.
RN -1- SUBUNIT: Homodimer.
RN -1- SUBCELLULAR LOCATION: Nuclear (Probable).
RN -1- ALTERNATIVE PRODUCTS:
RN Event=Alternative splicing; Named isoforms=2;
RN Comment=The level of each isoform is regulated by a splicing
RN event that occurs when the UPR is induced by IRE1;
RN Name=U;
RN IsoId=P41546-1; Sequence=Displayed;
RN Name=I;
RN IsoId=P41546-2; Sequence=VSP 000586;
RN Note=Active and stable isoform which induces UPR;
RN -1- SIMILARITY: Belongs to the bZIP family.
RN -1- SIMILARITY: Contains 1 bZIP domain.
RN This Swiss-Prot entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use as long as its content is in no way modified and this statement is not
RN removed.
RN EMBL; D26506; BAA05513.1; -; Genomic DNA.
RN EMBL; D50617; BAA24425.1; -; Genomic DNA.
RN EMBL; D86413; BAA19565.1; -; Genomic DNA.
RN PIR; S78571; S78571.
RN GeneOnline; 140124; -.
RN TRANSFAC; T02039; -.
RN Ensembl; YFL031W; Saccharomyces cerevisiae.
RN SGD; S00001863; HAC1
RN GO; GO:0005634; C:nucleus; TAS.
RN GO; GO:0003704; F:specific RNA polymerase II transcription fa. . .; TAS.
RN GO; GO:0003700; F:transcription factor activity; IDA.
RN GO; GO:0006644; P:phospholipid metabolism; TAS.
RN GO; GO:0045893; P:positive regulation of transcription, DNA-d. . .; IDA.
RN GO; GO:0006357; P:regulation of transcription from RNA polyme. . .; TAS.
RN GO; GO:0006990; P:unfolded protein response, positive regulat. . .; TAS.
RN InterPro; IPR011700; bZIP 2.
RN InterPro; IPR004827; TF_bZIP.
RN Pfam; PF07716; bZIP 2; 1.
RN SMART; SM00338; bZIP 1.
RN PROSITE; PS00217; bZIP; FALSE NEG.
RN PROSITE; PS00036; bZIP_BASIC; 1.
RN Activator; Alternative splicing; Complete proteome; DNA-binding;
RN Nuclear protein; Transcription; Transcription regulation;
RN Unfolded protein response.
FT DOMAIN 67 81 Leucine-zipper.
FT DNA_BIND 29 61 Basic motif.
FT VARSPLIC 221 230 AVITWRKIQ -> EAQSGLSFELNDFFITS (in
FT isoform 1).
FT FTID=VSP 000586.
FT RLCRCVRCFRVGRDFMGAAECLRRKMYQSRRLPYPTIN
FT NLFDAVASPLADLCDDIAGSLPFDNSIDLNDWRNPVIT
FT MTRKIQ -> ATLSPKSMRDSASDQSTSWELQMFKNVPE
FT STTLPAVDNNLEDAVASWQTHSATI (in Ref. 2).
FT RLCRCVRCFRVGRDFMGAAECLRRKMYQSRRLPYPTI
FT -> ATLSPKSMRDSASDQSTSWELQMFKNVPESTLLPAV
FT DN (in Ref. 4).
CONFLICT 143 230
CONFLICT 143 182

SQ SEQUENCE 230 AA; 26582 MW; 43073BCCCC4709B CRC64;
Query Match 11.7%; Score 206.5; DB 1; Length 230;
Best Local Similarity 40.3%; Pred. No. 0.0028;
Matches 52; Conservative 21; Mismatches 39; Indels 17; Gaps 3;
QY 66 LPVP---KTNLPKRAKTEDEKEQRIEVLNRRAAQAQTSRRKRLEMEKLESEKIDME 122
DB 16 LAITPFKSLTPPKRAKTKKEEQRRIELNRRAHQSRKKRLHLQLYLERKCSLLE 75
QY 123 Q-QNQFLLQRLAQWE-----AENRLSQVAQLSAEVRGSRHSTPTSSSPAS 168
DB 76 NLLNSVNLKLAHDHEDALTCSDAFVASLDEYRDFQSTRGASLDTRASSSSDTFTSP 135
QY 169 VSPFLTPTL 177
DB 136 LNCTMEPRL 144
RESULT 14
Q7ZYC2 XENLA
ID Q7ZYC2 XENLA PRELIMINARY; PRT; 396 AA.
AC 07ZYC2;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE xbp1-prov protein.
OS xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus; Xenopus.
OX NCBI_TaxID=8355;
RN NUCLEOTIDE SEQUENCE.
RP TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buotow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RN NUCLEOTIDE SEQUENCE.
RP TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative";
RL Dev. Dyn. 225:384-391(2002).
RN [3]
RN NUCLEOTIDE SEQUENCE.
RP TISSUE=Embryo;
RA Klein S., Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC043852; AAH43852.1; -; mRNA.
DR HSSP; P05412; 1JNM.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0005634; C:nucleus; IEA.

THIS PAGE BLANK (USPTO)

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	231.5	13.1	238	2	S78571	transcription fact
2	165.5	9.4	322	2	T08592	TGACG-motif-bindin
3	163.5	9.3	168	2	T05922	bzip protein Hys (
4	163	9.2	326	2	T08591	TGACG-motif bindin
5	162	9.2	322	2	T12093	TGACG-motif bindin
6	159.5	9.0	260	1	A36299	transcription fact
7	159	9.0	600	2	T00759	hypothetical prote
8	154.5	8.7	688	2	T32750	hypothetical prote
9	153.5	8.7	327	2	IS3459	Fra-2 - rat
10	143.5	8.1	1165	2	T16420	hypothetical prote
11	142.5	8.1	380	1	TVRFE5	transforming prote
12	141	8.0	323	2	A35909	fra-2 protein - ch
13	139.5	7.9	2364	2	A56577	microtubule-associ
14	138.5	7.8	1737	2	A59235	unconventional myo
15	137	7.8	403	2	T47621	bzip transcription
16	137	7.8	772	2	IS0463	protein kinase - c
17	137	7.8	1001	2	T16419	hypothetical prote
18	136.5	7.7	326	2	I48351	fos-related antige
19	136.5	7.7	826	2	T43638	casease-related pr
20	136	7.7	267	2	JC4857	hepatocarcinogenes
21	136	7.7	393	2	IS6321	light-induced prot
22	134.5	7.6	380	1	TVNSF	transforming prote
23	133	7.5	338	1	TVNSFB	transforming prote
24	133	7.5	381	1	TVMVJ	transforming prote
25	133	7.5	1380	2	T00365	hypothetical prote
26	133	7.5	1829	2	T24583	hypothetical prote
27	133	7.5	3938	2	T42761	Basoon protein -
28	132.5	7.5	777	2	F54024	protein kinase (EC
29	132.5	7.5	1621	2	A82255	hypothetical prote

```
QY 169 VSPPTPTTLFKQGEVPLDRIPFPPTPSVTDYSPTLKPSLSLAES----- 212
Db 136 LNCMEPA-----TILSPKSMRDSASDQETSWELQMFKTE 169
QY 213 --PDLTQHAPVSVGGLEGDESALTLPDLGASIKHEPTHDLTAPLSDDDPRRFLPENGDSLE 270
Db 170 NYPESTTLPAV-----DNN--NLFDVAVS-----PLADPLCDD-----IAGNS---- 205
QY 271 SDSSLLLEDGFAFDVLDSGLSAPFPDSDMYDED 302
Db 206 -----LPPDNSIDL 215

RESULT 2
T08592
TGACG-motif-binding protein STF2 - soybean
C:Species: Glycine max (soybean)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 31-Dec-2004
C:Accession: T08592
R:Cheong, Y.H.; Yoo, C.M.; Park, J.M.; Ryu, G.R.; Goekjian, V.H.; Nagao, R.T.; Key, J.L.
submitted to the EMBL Data Library, September 1995
A:Description: STF1 is a novel TGACG-binding factor with a zinc-finger motif and a bZIP
A:Reference number: Z16445
A:Accession: T08592
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-326 <CHE>
A:Cross-references: UNIPROT:Q39896; UNIPARC:UPI00000A61C8; EMBL:L28004; NID:g986966; PID
A:Experimental source: strain Williams; hypocotyl
C:Gene: STF2
C:Genetics:
C:Superfamily: TGACG-motif-binding transcription factor

Query Match 9.4%; Score 165.5; DB 2; Length 322;
Best Local Similarity 27.8%; Pred. No. 0.0065;
Matches 47; Conservative 31; Mismatches 66; Indels 25; Gaps 2;

QY 9 PVKMDAPANSPTTSPLEVPVLTSPADTSLRTKNVVAQTKEPKPAKRSWGQELPV 68
Db 158 PCATSDAQSMQTTISGQSEKVSLSVADPK-----QPGPESDEEIRRVPEIGESAG 209
QY 69 PKNLIP-----PRKRAKTEDEKQRIERVLRNRAAAQTSRERKRLEMEKIDMEQOQNFLLQRL 111
Db 210 TSASQPDAGSNAGTERTVQGTGQKGRSPADKESKRLKLLRNVSQAQAREKKAYL 269
QY 112 EKLESEKIDMEQOQNFLLQRLAQMENNRLSQOVAQLSAEVRGSRHST 160
Db 270 IDLETRVKDLEKKNSBELKRLTLQENQMLRQILKNTTASRRGSNNGT 318

RESULT 3
T50922
bZIP protein HYS [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004
C:Accession: T50922
R:Oyana, T.; Shimura, Y.; Okada, K.
submitted to the EMBL Data Library, July 1997
A:Description: The Arabidopsis HYS gene encodes a bZIP protein that regulates stimulus-i
A:Reference number: Z25271
A:Accession: T50922
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-168 <OYA>
A:Cross-references: UNIPROT:Q24646; UNIPARC:UPI000012CFQ4; EMBL:AB005295; PIDN:BA21116.
A:Experimental source: Landsberg erecta

Query Match 9.3%; Score 163.5; DB 2; Length 168;
Best Local Similarity 29.7%; Pred. No. 0.0039;
Matches 41; Conservative 28; Mismatches 68; Indels 1; Gaps 1;

QY 18 NSPTTSPLEVPVLTSPADTSLRTKNVVAQTKEPKPAKRSWGQELPVPKTNLP 77
Db 18 NSPTTSPLEVPVLTSPADTSLRTKNVVAQTKEPKPAKRSWGQELPVPKTNLP 77
```

```
Db 21 SSSAPHLEIKE-GIESDEEIRRVFPGGEAVGKETSREGSGSATGQERTQATVGSQRK 79
QY 78 RAKTEDEKQRIERVLRNRAAAQTSRERKRLEMEKIDMEQOQNFLLQRLAQMZA 137
Db 80 RQRTPAEKENKRLKLLRNVSQAQAREKKAYLSELENRVKDLNKNSELERLSTLQN 139
QY 138 ENNRLSQOVAQLSAEVRG 155
Db 140 ENQMLRHILKNTTGNKRG 157

RESULT 4
T08591
TGACG-motif binding protein STF1 - soybean
C:Species: Glycine max (soybean)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 31-Dec-2004
C:Accession: T08591
R:Cheong, Y.H.; Yoo, C.M.; Park, J.M.; Ryu, G.R.; Goekjian, V.H.; Nagao, R.T.; Key, J.L.
submitted to the EMBL Data Library, September 1995
A:Description: STF1 is a novel TGACG-binding factor with a zinc-finger motif and a bZIP
A:Reference number: Z16445
A:Accession: T08591
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-326 <CHE>
A:Cross-references: UNIPROT:Q39895; UNIPARC:UPI00000A6233; EMBL:L28003; NID:g2934883; PID
A:Experimental source: strain Williams; hypocotyl
C:Superfamily: TGACG-motif-binding transcription factor

Query Match 9.2%; Score 163; DB 2; Length 326;
Best Local Similarity 30.4%; Pred. No. 0.0093;
Matches 45; Conservative 28; Mismatches 53; Indels 22; Gaps 3;

QY 20 PTPPSLE-----VPVLTSPADTSLRTKNVVAQTKEPKPAKRSWGQELPVPKTN 72
Db 190 PKQPGPESDEEIRRVPEIGESAGTS-----ASRPDAGSNAGTGAQG-----TG 234
QY 73 LPPRKRKATEDEKQRIERVLRNRAAAQTSRERKRLEMEKIDMEQOQNFLLQRL 132
Db 235 DSQKGRSPADKESKRLKLLRNVSQAQAREKKAYLIDLETRVKDLEKKNSLKERL 294
QY 133 AQMEANNRLSQOVAQLSAEVRGSRHST 160
Db 295 STLQENQMLRQILKNTTASRRGSNGT 322

RESULT 5
T12093
TGACG-motif binding protein - fava bean
C:Species: Vicia faba (fava bean)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 31-Dec-2004
C:Accession: T12093
R:Wohlfarth, T.
submitted to the EMBL Data Library, May 1996
A:Reference number: Z17415
A:Accession: T12093
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-322 <WOH>
A:Cross-references: UNIPROT:O04234; UNIPARC:UPI00000A297E; EMBL:X97904
C:Superfamily: TGACG-motif-binding transcription factor

Query Match 9.2%; Score 162; DB 2; Length 322;
Best Local Similarity 29.7%; Pred. No. 0.01;
Matches 44; Conservative 27; Mismatches 55; Indels 22; Gaps 3;

QY 20 PTPPSLE-----VPVLTSPADTSLRTKNVVAQTKEPKPAKRSWGQELPVPKTN 72
Db 186 PKQPGLESDEEMRRVPDNGGESAGTSASHKGTGSTAGPERAQGTGE---GQ----- 233
QY 73 LPPRKRKATEDEKQRIERVLRNRAAAQTSRERKRLEMEKIDMEQOQNFLLQRL 132
Db 234 ---KKRGRSPADKESKRLKLLRNVSQAQAREKKAYLSDLETRVNDLEKKNSLKERL 290
```


QY 133 AQMEAEENRLSQVAQLSAEVRGRSHST 160
Db 291 STLQENQMLRQLKNTTASRGNGSGT 318

RESULT 6
A36299
transcription factor hXBP-1 - human
N;Alternate names: DNA-binding protein TREBS
C;Species: Homo sapiens (man)
C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: A36299, S12559, A47490
R;Liou, H.C.; Boothby, M.R.; Finn, P.W.; Davidson, R.; Nabavi, N.; Zeleznik-Le, N.J.; Tir
Science 247, 1581-1583, 1990
A;Title: A new member of the leucine zipper class of proteins that binds to the HLA DR α
A;Reference number: A36299; MUID:90208323; PMID:2321018
A;Accession: A36299
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-260 <L10>
A;Cross-references: UNIPROT:P17861; UNIPARC:UPI0000031C60; GB:M31627; NID:g184485; PIDN:
R;Yoshimura, T.; Fujisawa, J.I.; Yoshida, M.
EMBO J. 9, 2537-2542, 1990
A;Title: Multiple cDNA clones encoding nuclear proteins that bind to the tax-dependent e
A;Reference number: S12559; MUID:90316112; PMID:2196176
A;Accession: S12559
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-32, 'GQA', 35-260 <YOS>
A;Cross-references: UNIPARC:UPI00006FC05; EMBL:X55543; NID:g287644; PIDN:CAA39149.1; PI
R;Ponath, P.D.; Fass, D.; Liou, H.C.; Glimcher, L.H.; Strominger, J.L.
J. Biol. Chem. 268, 17074-17082, 1993
A;Title: The regulatory gene, hXBP-1, and its target, HLA-DRA, utilize both common and d
A;Reference number: A47490; MUID:93352484; PMID:8349596
A;Accession: A47490
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-32, 'GQA', 35-128, 'T', 130-192, 'F', 194-198, 'R', 200-260 <PON>
A;Cross-references: UNIPARC:UPI0000173305; GB:L13850
A;Note: authors translated the codon ACC for residue 130 as Asn, AAG for residue 151 as
C;Genetics:
A;Gene: GDB:XBP1; XBP2
A;Cross-references: GDB:131393; OMIM:194355
A;Map position: 22pter-22qter
A;Intons: 76/3; 108/3; 151/3; 200/3
C;Superfamily: X box-binding protein 1; fos/jun DNA-binding domain homology
C;Keywords: DNA binding; leucine zipper; nucleus; transcription regulation
F;64-104/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 9.0%; Score 159.5; DB 1; Length 260;
Best Local Similarity 28.8%; Pred. No. 0.011;
Matches 55; Conservative 31; Mismatches 80; Indels 25; Gaps 6;

QY 15 AFANSTPTSPLEVPVLTVSADTSRLTKNVVAQTKEPKPAKKRKSQGLPVPKTNLP 74
Db 5 AAAPNPADGTPFKVLLISQGPAS-----AGAPAAFLPMPVPAQRGASPEAASGGLP 55

QY 75 -PKRAK-TEDEKEQRRIEVLNRNAAQTSRERKLEMEKLESEKIDMEQQQFLT--- 129
Db 56 QAKRKQRLTHLSPEEKALRKLNVAQTARDKKAWSLEQQVVDLEENQKLLLEN 115

QY 130 ----QRLAQMEAEENRLSQ-----VAQLSAEVRGRSHSTPTSSPASVSPILTPTLTK 179
Db 116 QLLREKTHGLVWENQELRQLGMDALVAEEAEAKGN-EVRPVAGSAESAALRLAPLQ 174

QY 180 QEGDEVPDLRI 190
Db 175 VQAQLSPLQNI 185

RESULT 7
T00759
hypothetical protein At2g40950 [imported] - Arabidopsis thaliana

N;Alternate names: hypothetical protein T20B5.15
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 31-Dec-2004
C;Accession: T00759; H84835
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul,
submitted to the EMBL Data Library, November 1997
A;Description: Arabidopsis thaliana chromosome II BAC T20B5 genomic sequence.
A;Reference number: Z14159
A;Accession: T00759
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-600 <R0U>
A;Cross-references: UNIPROT:Q22208; UNIPARC:UPI0000179D30; EMBL:AC002409; NID:g2623294; I
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; L
m.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
Eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: H84835
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-600 <STO>
A;Cross-references: UNIPARC:UPI0000179D30; GB:AE002093; NID:g2623309; PIDN:AAB86455.1; GS
C;Genetics:
A;Gene: T20B5.15; At2g40950
A;Map position: 2
C;Superfamily: bZIP transcription factor; fos/jun DNA-binding domain homology

Query Match 9.0%; Score 159; DB 2; Length 600;
Best Local Similarity 25.8%; Pred. No. 0.033;
Matches 70; Conservative 46; Mismatches 109; Indels 46; Gaps 9;

QY 5 DRFPVPMEDAFANSTPTSPLEVPVLT-----VSPA--DTSRLTKNVVAQTKEPK 53
Db 123 NRESPRDSDRCGADHNLDTPLSSQSGSGCGSDSEATNESPFSRNVAVDQVKVE 182

QY 54 KPA-----KKRKSQGLPVPKTNLPKRAKTED-----EKEQRRIEVLNRNAA 100
Db 183 EAATTTTITKRKKEIDELTDESNRNKKYRSRGSDADASAVTGEDEKKRLMRNESA 242

QY 101 QTSREKRLEMEKLESEKIDMEQQQFLQLQRLAQMEAEENRLSQVAQLSAEVRGRSHST 160
Db 243 QLSRQKQKHVVELEEKVRNMHSTITDLNGKISYFAENATLRQ---QLGGNGMCPHLP 299

QY 161 PTSSSPASVSPILTPT-----LFKQEGDEVPDLRIPTFPPTSPVTDYSPTLKPSSLA 210
Db 300 P---PMGMYPMPAMPYPMPCPPVMVQSQSQVPLIPRLKPQNTLTGTSKAKKSEK 356

QY 211 ESPDLTQHPA-VSVGGLGEGESALTFLDIGA 240
Db 357 KSEAKTKVASISFLGL-----LFCILFLGA 382

RESULT 8
T32750
hypothetical protein F57B10.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C;Accession: T32750
R;Greco, T.; Elliott, G.; Keppler, D.
submitted to the EMBL Data Library, December 1997
A;Description: The sequence of C. elegans cosmid F57B10.
A;Reference number: Z21219
A;Accession: T32750
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-688 <GRE>
A;Cross-references: UNIPROT:Q44743; UNIPARC:UPI0000079D06; EMBL:AF039713; PIDN:AAB96719.1
A;Experimental source: strain Bristol N2; clone F57B10
C;Genetics:
A;Gene: CESP:F57B10.1

A:Map position: 1
A:Introns: 54/1; 150/3; 196/1; 306/3; 383/3; 627/3

Query Match 8.7%; Score 154.5; DB 2; Length 688;
Best Local Similarity 25.7%; Pred. No. 0.073;
Matches 49; Conservative 41; Mismatches 74; Indels 27; Gaps 6;

QY 19 SPTTSPLEVPVLTSPADTSLRTKNVVAQTKEPKPAKKRSWGQELPVPKTNLPNPPR-- 76
DB 204 NPASISLNAPSSSFNPQST--STATSSSSSSSTNGGFKVSTGERRKYPLRLDEEEI 261

QY 77 KRAKTE-----DEKEQRRIERVLNRAAAQTSRRKRKLEMEKLESEKIDMEQ 123
DB 262 KLCKKEGICLPDFFPLTKAEERDLKIRRKIRKRSQAQTSRRKKQDYIQLEDREVSESTK 321

QY 124 QNOFLORLAQMEANRLSQVAQLSAEV-RGSRHSTPTS-----SSPASVSPITL 174
DB 322 ENQALKQQIERLSSENQSVISQLKLAQLGQNAKRTTQAGRCLAVFMLSACLLVSPQLS 381

QY 175 PTLFKQEGDEV 185
DB 382 P-LGNQDNQKV 391

RESULT 9
Fra-2 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
C:Accession: I55459
R:Balser, R.D.; Klein, D.C.
J. Biol. Chem. 270, 27319-27325, 1995
A:Title: Circadian Expression of Transcription Factor Fra-2 in the Rat Pineal Gland.
A:Reference number: I55459; MUID:96070844; PMID:7592294
A:Accession: I55459
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-327 <RES>
A:Cross-references: UNIPROT:P51145; UNIPARC:UPI000012ABD0; EMBL:U18913; NID:g1001950; PI
C:Genetics:
A:Gene: Fra-2
C:Superfamily: fos transforming protein; fos/jun DNA-binding domain homology
F:120-160/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 8.7%; Score 153.5; DB 2; Length 327;
Best Local Similarity 24.2%; Pred. No. 0.034;
Matches 82; Conservative 43; Mismatches 121; Indels 93; Gaps 15;

QY 18 NSPTTSPLEVPVLTSPADTSLRTKNVVAQTKEPKPAKKRSWGQELPVP-----KT 71
DB 52 NAITTTSQDLQWVQPTVITSM--SNPYPRSHPSPLGLRSVPGHMALPRPGVIKTIGT 109

QY 72 NLPPRKAK--TEDEKEQRRIERVLNRAAAQTSRRKRKLEMEKLESEKIDMEQNOFL 129
DB 110 TVGRRRDEQLSPREEEKRIIR--RNKLAARACNRNRRELTEKLTQTEEELEEKSGIQ 168

QY 130 ORLAQMEANRLS-----QVAQLSAEVGRGSRHSTPTS--SSPASVSPITLPTLTKQE 181
DB 169 KGIAELQKEKEKLEFVLVAGHPVCKISPE---ERRSPISGVQSLRGTSAGVGPVVVKQE 225

QY 182 GDEVPDLRIFFPTTSPVDTSPYLKPSLSLAESPDLTQHPAVSVGGLEGDESALTFLDLGAS 241
DB 226 -----PPED-----SPSSA-GMDKTQRSVIKPSIAG-----GGF 256

QY 242 IKHEPTHDLTAPLSDDDFRRLFNGDSLSLSDSLLEDGFADFVLDGDLSPAPFDSMVD 301
DB 257 YGEEPLH---TPFVVTSTPAITPGTNSL-----VFTYPSV----- 288

QY 302 DTEPVTLEDLEOTNGLSDSASCAASLQPSHGASTSRCD 340
DB 289 -----LEQESPASPSESCKA-----HRRSSSGD 313

RESULT 10
Ti6420
hypothetical protein F52C9.8b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: Ti6420
R:Favella, T.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid F52C9.
A:Reference number: Z18511
A:Accession: Ti6420
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1165 <FAV>
A:Cross-references: UNIPARC:UPI000017B9F5; EMBL:U39850; NID:g1055052; PID:g1055056; PIDN:
C:Genetics:
A:Gene: CESP:F52C9.8b
A:Introns: 21/1; 36/1; 61/3; 79/3; 107/3; 464/3; 566/1; 1010/3; 1029/3; 1106/3

Query Match 8.1%; Score 143.5; DB 2; Length 1165;
Best Local Similarity 23.1%; Pred. No. 0.62;
Matches 87; Conservative 57; Mismatches 140; Indels 93; Gaps 18;

QY 3 SADRFSPVKMEDAFANSTPTTSPLEVPVLTSPADTSLRTKNVVAQTKEPKPAKKRSW 62
DB 528 SESKIEPVDVKPRVA--PVPP--QVPVTPTKPVITNNKKRIDVVVTLDED---APRRVQV 580

QY 63 GOELPVPKTNLPKRAKTEDEKEQRRIERVLNRAAAQTSR-ERKRLMEKLESE-KID 120
DB 581 KQEIPEVSTSDATKSDAAPTARGAVRIKQEVESDVAPNTILISAKKFERMKAEDKED 640

QY 121 ME-----QONQFLQLRLAQMEAE-----NNRLSQ-----QVAQLS---AEVGRSR 157
DB 641 MKKKTAAALQEALFNIEQERRRVEKEIAAFATTNQAQVPQNPASSVQIAQVSTSESAPGTS 700

QY 158 H--STPTSSSPASVSPITLPTLPKQEGDEVPLDLRPPF-----TPSVTDYSPTLKP 206
DB 701 EAAATETMTSPKTNVIVETSGEOEDE--DEIPIKKKRRAKIVNSDBEEEPVRHP 757

QY 207 SSLAESPDLTQHPAVSVGGLEGDESALTFLDLGASIK-----HEPTHDLTAPLSDDDF 259
DB 758 KRSDEKEKRH-----VSVAESDD-----DMPVVKKRRNQSPEDPEVSASPSDEDE- 806

QY 260 RRLFNGDSLSLSDSLLEDGFADFVLDGDLSPAPFDSMVDPTPTLTDLEQTNGL--- 317
DB 807 -----DDIGSFVVSNDNDDA---DSFVVGDESPIEYEEDDDMIER 846

QY 318 -----SDSASCKAAS 327
DB 847 RSSRKRSDRSRKSAT 863

RESULT 11
TVRTFS
transforming protein fos - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 09-Jul-2004
C:Accession: A28263; I59159
R:Curran, T.; Gordon, M.B.; Rubino, K.L.; Sambucetti, L.C.
Oncogene 2, 79-84, 1987
A:Title: Isolation and characterization of the c-fos(rat) cDNA and analysis of post-transcriptional regulation of the c-fos gene.
A:Reference number: A28263; MUID:88143713; PMID:3325886
A:Accession: A28263
A:Molecule type: mRNA
A:Residues: 1-380 <CUR>
A:Cross-references: UNIPROT:P12841; UNIPARC:UPI000012AB6E; GB:X06769; NID:g55933; PIDN:CJ
R:Abate, C.; Luk, D.; Gentz, R.; Rauscher, F.J.
Proc. Natl. Acad. Sci. U.S.A. 87, 1032-1036, 1990
A:Title: Expression and purification of the leucine zipper and DNA-binding domains of Fos.
A:Reference number: I59159; MUID:90138931; PMID:2105492
A:Accession: I59159
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA

C:Genetics:
A:Gene: myom
A:Map position: 6, aldB-cabA2
F:62-874/Domain: myosin motor domain homology #status atypical <MMO>

Query Match 7.8%; Score 138.5; DB 2; Length 1737;
Best Local Similarity 25.3%; Pred. NO. 2;
Matches 60; Conservative 37; Mismatches 91; Indels 49; Gaps 8;

QY 40 RTQNVVATQKPEEKPAKKRKSQGLPVPKTNLPFRKAKTEDEKQRRRIERVLRPA 99
DB 961 RQKELQRREERKELEKKEEREL-----ERQKEKEKQE-----RKRKE 1004
QY 100 AQTSRERKLEMEKLESKIDMEQOQFLQRLAQMEANNRLSQOVAQLSAEVRGSRHS 159
DB 1005 BEKQERKKKEKEKEKKEKKK-----KNEQLSLPSLDITNSPSL 1050
QY 160 TPTSSPASVSPTLTPTLFQEGDEVLDRIPTFPPTPSVTDYSPTLKPSLSLAESPDLTQHP 219
DB 1051 INTTTTTTTTTTTNT-----SSPPLSPPTSPRPST-----PSTSSSSSTTSSP 1096
QY 220 AVSVGGLGEGDESALTFLDLGASIKHEPTHDLTAPLSDDDFRFLFNGDSS--LESDDS 274
DB 1097 STKKQLLFKFNISNL--LSKSL-HGSHSDKNSKEDNNSNNNGDSTIILSSDSS 1150

RESULT 15
T47621
bZIP transcription factor-like protein - Arabidopsis thaliana
N:Alternate names: protein T14E10.190
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 31-Dec-2004
C:Accession: T47621
R:Obermaier, B.; Ottenwaelder, B.; Duchemin, D.; Zeitler, K.; Mewes, H.W.; Lemcke, K.; M
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24463
A:Accession: T47621
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-403 <OBE>
A:Cross-references: UNIPROT:Q9MIG6; UNIPARC:UPI00000A86DF; EMBL:AL1138656
A:Experimental source: cultivar Columbia; BAC clone T14E10
C:Genetics:
A:Map position: 3
A:Introns: 152/3; 173/1; 235/2; 260/3; 302/3
A>Note: T14E10.190
C:Superfamily: BZIP protein; fos/jun DNA-binding domain homology

Query Match 7.8%; Score 137; DB 2; Length 403;
Best Local Similarity 23.7%; Pred. No. 0.41;
Matches 54; Conservative 36; Mismatches 86; Indels 52; Gaps 8;

QY 31 TVSPADTSLRTKN-----VVAQTKP-----EEKPAKKRKSQGLPVPKT 71
DB 155 TVKPEDSSASASQKQAQSIQAQTSFGASSVRFSTTSTQKKP-----DVPARQT 205
QY 72 NLPPRK-----RAKTEDEKEQRIERVLRNRAAQTSRERKLEMEKLESKIDME 122
DB 206 SISRDDDDDLGDGADNGDPTVKRARRMLNSRESARRRRKQEQNNEFDTVQGQLR 265
QY 123 QCNQFLQRLAQME-----AENRLSQOVAQLSAEVRGSRHSTPTSSPASVSPTLTPT 176
DB 266 AEHSTLNLSDMNHKYDAAAVDNRI-----LRADIETLRTKVKMAEETVKRVGTGNFL 319
QY 177 LFKQEGDEVPLDRIPFPPTPSVTDYS--PTLKP--SSLAESPDLTQHPAV 221
DB 320 HWSRPNMGIPFNTSASSIIPNSNHILKPANSSTNTSAGLAQNRV 367

Search completed: November 23, 2005, 03:26:04
Job time : 26.5265 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:04:00 ; Search time 136.303 Seconds
(without alignments)
1125.017 Million cell updates/sec

Title: US-10-663-450-4
Perfect score: 1766
Sequence: 1 MKSADRFSPVKMEDAFANSP.....PSHGASTSRCDGGIAGSA 349

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*
9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1766	100.0	349	8	AdS12805 Aspergill
2	1766	100.0	349	8	AdS12803 Aspergill
3	1766	100.0	349	8	AdS12863 Aspergill
4	1755.5	99.4	350	4	AaB82976 Aspergill
5	1755.5	99.4	350	5	AaE15372 Aspergill
6	1293.5	73.2	342	4	AaB82977 Aspergill
7	1293.5	73.2	342	5	AaE15381 Aspergill
8	1293.5	73.2	342	5	AdS12818 Aspergill
9	1293.5	73.2	386	5	AaE15379 Aspergill
10	1293.5	73.2	386	8	AdS12815 Aspergill
11	434	24.6	451	5	AaE15371 Trichoder
12	434	24.6	451	8	AdS12801 Trichoder
13	434	24.6	451	8	AdS12804 Trichoder
14	419.5	23.8	450	4	AaB82975 Trichoder
15	398.5	22.6	409	7	ABO43144 A. thalia
16	398.5	22.6	409	7	AdB31925 Plant (A.
17	398.5	22.6	409	8	AdO2271 Thalacro
18	325	18.4	64	5	AaE15374 Aspergill
19	288	16.3	64	5	AaE15373 Trichoder
20	231.5	13.1	230	2	AAW53806 Transcrip
21	231.5	13.1	230	8	AdT87049 Yeast Str
22	231.5	13.1	238	2	AAW53807 Transcrip
23	216	12.2	200	8	AdS43437 Bacteri
24	192	10.9	376	7	Adf18647 Human spl

25	192	10.9	376	8	ADN16844	Adn16844 Human XB1
26	192	10.9	376	9	AD266508	Ad266508 Amino aci
27	192	10.9	376	9	AEb98685	Aeb98685 Endoplasm
28	192	10.9	928	9	ADY71599	Ady71599 Human IRE
29	178	10.1	84	5	ABP02534	Abp02534 Human ORF
30	176.5	10.0	68	5	AAe15382	AAe15382 Yeast HAC
31	176.5	10.0	68	8	ADs12859	AdS12859 Saccharom
32	164.5	9.3	634	8	ADP98908	Adp98908 C. albica
33	164	9.3	370	9	ADV86805	Adv86805 Transcrip
34	163.5	9.3	168	5	AAU93013	Aau93013 Arabidops
35	163.5	9.3	168	7	ADD30174	Add30174 Plant yie
36	163.5	9.3	168	8	ADi43893	Adi43893 Plant tra
37	163.5	9.3	211	3	AAG08861	Aag08861 Arabidops
38	160	9.1	261	6	ABR47631	Abr47631 Breast ca
39	160	9.1	261	7	ADb75629	Adb75629 Prostate
40	160	9.1	261	8	ADL83138	Adl83138 Human PRO
41	160	9.1	261	8	ADN16846	Adn16846 Human XB1
42	160	9.1	261	8	ADR14692	Adr14692 Human NF-
43	160	9.1	261	8	ABM82329	Abm82329 Tumour-as
44	160	9.1	261	9	ADX07539	Adx07539 Cyclin-de
45	160	9.1	261	9	ADY71598	Ady71598 Human non

ALIGNMENTS

RESULT 1
ADs12805
ID ADs12805 standard; protein; 349 AA.
XX ADs12805;
AC ADs12805;
XX
DT 16-DEC-2004 (first entry)
XX
DE Aspergillus nidulans hacA DNA binding domain.
XX
KW unfolded protein response; UPR; HAC1; PTC2;
KW unfolded protein response modulator; enzyme production; hacA; chaperone;
KW foldase; DNA binding domain.
XX
OS Emericella nidulans.
XX
PN US2004186070-A1.
XX
PD 23-SEP-2004.
XX
PF 15-SEP-2003; 2003US-00663450.
XX
PR 24-MAR-2000; 2000US-00534692.
PR 23-MAR-2001; 2001US-00816277.
(GEMV) GENENCOR INT INC.
PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX WPI; 2004-707924/69.
DR
PT Increasing the secretion of a heterologous protein, such as a therapeutic
PT or an industrial enzyme, in genetically modified eukaryotic cells by
PT inducing an elevated unfolded protein response (UPR).
XX
XX Example 3; SEQ ID NO 6; 83pp; English.
The invention describes a method of increasing the secretion of a
heterologous protein in a eukaryotic cell, comprising inducing an
elevated unfolded protein response (UPR). Also described are: an isolated
nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
unfolded protein response and has less than 50% similarity to yeast HAC1
protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
protein induces unfolded protein response and where the HAC1 protein
comprises a DNA binding region that has greater than 70% similarity to
the DNA binding region of filamentous fungi HAC1 protein; a protein
having unfolded protein response inducing activity and having greater

CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
CC respectively) as given in the specification; a protein having an amino
CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
CC response and has at least 70% similarity to a fully defined amino acid
CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
CC specification; an isolated nucleic acid encoding a IRE1 protein that
CC modulates unfolded protein response and has at least 60% similarity to a
CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
CC having unfolded protein response modulating activity and having greater
CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
CC heterologous nucleic acid encoding a protein having unfolded protein
CC response modulating activity and a heterologous nucleic acid encoding a
CC protein of interest to be secreted. The methods and compositions of
CC genetically manipulating cells to have an elevated unfolded protein
CC response (UPR) resulting in an increased capacity to produce secreted
CC proteins, are useful in e.g. production of therapeutic or industrial
CC enzymes. This is the amino acid sequence of *Aspergillus nidulans* HAC1
CC chaperone and foldase DNA binding domain.
XX
SQ Sequence 349 AA;

Query Match 100.0%; Score 1766; DB 8; Length 349;
Best Local Similarity 100.0%; Pred. No. 8.9e-140;
Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKSADRFSPVKMEDAFANSPPTPSLEVPVLTSPADTSIRTKNVVAQTPEEKPAKKRK 60
DB 1 MKSADRFSPVKMEDAFANSPPTPSLEVPVLTSPADTSIRTKNVVAQTPEEKPAKKRK 60
QY 61 SWGQELPVPKTNLPPRKRAKTEDEKEQRRIERVLNRNAAQTSSRERKRLMEKLESEKID 120
DB 61 SWGQELPVPKTNLPPRKRAKTEDEKEQRRIERVLNRNAAQTSSRERKRLMEKLESEKID 120
QY 121 MEQQNQFLQLRLAQMEAEENNRLSQQVAQLSAEVRGSRHSTPTSSSPASVSPTLTPTLFKQ 180
DB 121 MEQQNQFLQLRLAQMEAEENNRLSQQVAQLSAEVRGSRHSTPTSSSPASVSPTLTPTLFKQ 180
QY 181 EGDEVPLDRIPPTPSVTDYSPTLKPSLSAESPDLTQHHPAVSVGGLEGDESALTFLDLGA 240
DB 181 EGDEVPLDRIPPTPSVTDYSPTLKPSLSAESPDLTQHHPAVSVGGLEGDESALTFLDLGA 240
QY 241 SIKHEPTHDLTAPLSDDDFRRLFNCGDSSLSLESDDSGDGFADVLDSGDLSPFPDMSWD 300
DB 241 SIKHEPTHDLTAPLSDDDFRRLFNCGDSSLSLESDDSGDGFADVLDSGDLSPFPDMSWD 300
QY 301 FDEPVTLEDLQTNGLSDSASCKAASLQPSHGASTSRCDGGGIAAGSA 349
DB 301 FDEPVTLEDLQTNGLSDSASCKAASLQPSHGASTSRCDGGGIAAGSA 349

RESULT 2
ADSI2803
ID ADSI2803 standard; protein; 349 AA.

XX ADSI2803;
XX
DT 16-DEC-2004 (first entry)
XX
DE *Aspergillus nidulans* hacA chaperone and foldase #1.
XX unfolded protein response; UPR; HAC1; PTC2;
KW unfolded protein response modulator; enzyme production; hacA; chaperone;
KW foldase; enzyme.
XX
OS *Emicella nidulans*.
XX
PN US2004186070-A1.
XX
PD 23-SEP-2004.
XX
PF 15-SEP-2003; 2003US-00663450.

XX
PR 24-MAR-2000; 2000US-00534692.
PR 23-MAR-2001; 2001US-00816277.
XX
XX (GEV) GENENCOOR INT INC.
XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
PI WPI: 2004-707924/59.
DR N-PADB; ADSI2802.
XX
XX Increasing the secretion of a heterologous protein, such as a therapeutic
PT or an industrial enzyme, in genetically modified eukaryotic cells by
PT inducing an elevated unfolded protein response (UPR).

PS Example 3; SEQ ID NO 4; 83pp; English.

XX The invention describes a method of increasing the secretion of a
CC heterologous protein in a eukaryotic cell, comprising inducing an
CC elevated unfolded protein response (UPR). Also described are: an isolated
CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
CC unfolded protein response and has less than 50% similarity to yeast HAC1
CC protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
CC protein induces unfolded protein response and where the HAC1 protein
CC comprises a DNA binding region that has greater than 70% similarity to
CC the DNA binding region of filamentous fungi HAC1 protein; a protein
CC having unfolded protein response inducing activity and having greater
CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
CC respectively) as given in the specification; a protein having an amino
CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
CC response and has at least 70% similarity to a fully defined amino acid
CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
CC specification; an isolated nucleic acid encoding a IRE1 protein that
CC modulates unfolded protein response and has at least 60% similarity to a
CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
CC having unfolded protein response modulating activity and having greater
CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
CC heterologous nucleic acid encoding a protein having unfolded protein
CC response modulating activity and a heterologous nucleic acid encoding a
CC protein of interest to be secreted. The methods and compositions of
CC genetically manipulating cells to have an elevated unfolded protein
CC response (UPR) resulting in an increased capacity to produce secreted
CC proteins, are useful in e.g. production of therapeutic or industrial
CC enzymes. This is the amino acid sequence of *Aspergillus nidulans* HAC1
CC chaperone and foldase. Note: This sequence differs from ADSI2863 given in
XX figure 8.

SQ Sequence 349 AA;

Query Match 100.0%; Score 1766; DB 8; Length 349;
Best Local Similarity 100.0%; Pred. No. 8.9e-140;
Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKSADRFSPVKMEDAFANSPPTPSLEVPVLTSPADTSIRTKNVVAQTPEEKPAKKRK 60
DB 1 MKSADRFSPVKMEDAFANSPPTPSLEVPVLTSPADTSIRTKNVVAQTPEEKPAKKRK 60
QY 61 SWGQELPVPKTNLPPRKRAKTEDEKEQRRIERVLNRNAAQTSSRERKRLMEKLESEKID 120
DB 61 SWGQELPVPKTNLPPRKRAKTEDEKEQRRIERVLNRNAAQTSSRERKRLMEKLESEKID 120
QY 121 MEQQNQFLQLRLAQMEAEENNRLSQQVAQLSAEVRGSRHSTPTSSSPASVSPTLTPTLFKQ 180
DB 121 MEQQNQFLQLRLAQMEAEENNRLSQQVAQLSAEVRGSRHSTPTSSSPASVSPTLTPTLFKQ 180
QY 181 EGDEVPLDRIPPTPSVTDYSPTLKPSLSAESPDLTQHHPAVSVGGLEGDESALTFLDLGA 240
DB 181 EGDEVPLDRIPPTPSVTDYSPTLKPSLSAESPDLTQHHPAVSVGGLEGDESALTFLDLGA 240
QY 241 SIKHEPTHDLTAPLSDDDFRRLFNCGDSSLSLESDDSGDGFADVLDSGDLSPFPDMSWD 300
DB 241 SIKHEPTHDLTAPLSDDDFRRLFNCGDSSLSLESDDSGDGFADVLDSGDLSPFPDMSWD 300

Db 241 SIKHEPTHDLTAPLSDDDFRRLFNGDSSLSLSDSLLEDGFAFDVLDGSLDLSAPFPDSMVD 300
QY 301 FDTPEVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDGGQIAAGSA 349
DB 301 FDTPEVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDGGQIAAGSA 349

RESULT 3
ADSI12863
ID ADSI12863 standard; protein; 349 AA.
AC ADSI12863;
XX
XX
DT 16-DEC-2004 (first entry)
DE Aspergillus nidulans hacA chaperone and foldase #2.
KW unfolded protein response; UPR; HAC1; PTC2;
KW unfolded protein response modulator; enzyme production; HAC1; chaperone;
KW foldase; enzyme.
XX
OS Emericella nidulans.
XX
XX US2004186070-A1.
XX
XX 23-SEP-2004.
XX
XX 15-SEP-2003; 2003US-00663450.
XX
XX 24-MAR-2000; 2000US-00534692.
PR 23-MAR-2001; 2001US-00816277.
XX
XX (GEMV) GENENCOR INT INC.
XX
XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX WPI; 2004-707924/69.
DR N-PSDB; ADSI12802.
XX
XX
PT Increasing the secretion of a heterologous protein, such as a therapeutic
PT or an industrial enzyme, in genetically modified eukaryotic cells by
PT inducing an elevated unfolded protein response (UPR).
XX
PS Example 3; Fig 8; 83pp; English.

CC The invention describes a method of increasing the secretion of a
CC heterologous protein in a eukaryotic cell, comprising inducing an
CC elevated unfolded protein response (UPR). Also described are: an isolated
CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
CC unfolded protein response and has less than 50% similarity to yeast HAC1
CC protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
CC protein induces unfolded protein response and where the HAC1 protein
CC comprises a DNA binding region that has greater than 70% similarity to
CC the DNA binding region of filamentous fungi HAC1 protein; a protein
CC having unfolded protein response inducing activity and having greater
CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
CC respectively) as given in the specification; a protein having an amino
CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
CC response and has at least 70% similarity to a fully defined amino acid
CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
CC specification; an isolated nucleic acid encoding a IRE1 protein that
CC modulates unfolded protein response and has at least 60% similarity to a
CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
CC having unfolded protein response modulating activity and having greater
CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
CC heterologous nucleic acid encoding a protein having unfolded protein
CC response modulating activity and a heterologous nucleic acid encoding a
CC protein of interest to be secreted. The methods and compositions of
CC genetically manipulating cells to have an elevated unfolded protein
CC response (UPR) resulting in an increased capacity to produce secreted
CC proteins, are useful in e.g. production of therapeutic or industrial

CC enzymes. This is the amino acid sequence of Aspergillus nidulans HAC1
CC chaperone and foldase. Note: This sequence differs from ADSI12803 given in
CC the sequence listing.
XX
SQ Sequence 349 AA;
Query Match 100.0%; Score 1766; DB 8; Length 349;
Best Local Similarity 100.0%; Pred. No. 8.9e-140;
Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKSADRFSPVKMEDAFANSPPTTPSLSEVPVLTVPADTSLRTKNVVAQTKEPKKPAKRRK 60
DB 1 MKSADRFSPVKMEDAFANSPPTTPSLSEVPVLTVPADTSLRTKNVVAQTKEPKKPAKRRK 60
QY 61 SWGQELPVKTNLPPRKRAKTEDEKEQRRIERNRNRNAAQTSRERKRLMEKLESEKID 120
DB 61 SWGQELPVKTNLPPRKRAKTEDEKEQRRIERNRNRNAAQTSRERKRLMEKLESEKID 120
QY 121 MEQONQFLRLAQMEEAENNRSLSQVAQLSAEVRGSRHSTPTSSSPASVPTLTPTLFKQ 180
DB 121 MEQONQFLRLAQMEEAENNRSLSQVAQLSAEVRGSRHSTPTSSSPASVPTLTPTLFKQ 180
QY 181 EGDEVPLDRIPPTPSVTDSPTLKPSSLAESPDLTQHPAVSVGGLEGDESALTFLDLGA 240
DB 181 EGDEVPLDRIPPTPSVTDSPTLKPSSLAESPDLTQHPAVSVGGLEGDESALTFLDLGA 240
QY 241 SIKHEPTHDLTAPLSDDDFRRLFNGDSSLSLSDSLLEDGFAFDVLDGSLDLSAPFPDSMVD 300
DB 241 SIKHEPTHDLTAPLSDDDFRRLFNGDSSLSLSDSLLEDGFAFDVLDGSLDLSAPFPDSMVD 300
QY 301 FDTPEVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDGGQIAAGSA 349
DB 301 FDTPEVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDGGQIAAGSA 349

RESULT 4
AAB82976
ID AAB82976 standard; protein; 350 AA.
XX
AC AAB82976;
XX
DT 11-SEP-2003 (revised)
DT 21-DEC-2001 (first entry)
XX
DE Aspergillus nidulans hacA, involved in unfolded protein response.
XX
KW HAC1; transcription factor; unfolded protein response; protein secretion.
OS Emericella nidulans.
FH Key Location/Qualifiers
FT Domain 53..116
FT /label= DNA binding domain
XX
PN WO200172783-A2.
XX
XX 04-OCT-2001.
PD
PD 23-MAR-2001; 2001WO-US009401.
XX
XX 24-MAR-2000; 2000US-00534692.
XX
PA (GEMV) GENENCOR INT INC.
XX
PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX
DR WPI; 2001-626252/72.
DR N-PSDB; AAH26932.
XX
PT Increasing secretion of heterologous proteins e.g. lipase and cellulase
PT in eukaryotic cells useful in industry to increase production and
PT facilitate purification, by inducing an elevated unfolded protein
PT response.

XX Claim 54; Fig 8A-B; 89pp; English.

PS The present sequence is that of the hacA protein of Aspergillus nidulans,

XX as deduced from the newly isolated hacA gene (see AHA2932). HacA protein

CC is a transcription factor involved in the unfolded protein response

CC (UPR). The invention provides methods for increasing the secretion of a

CC heterologous protein in a cell by inducing an elevated UPR. This can be

CC achieved by modulating the activity of HAC1 (or hacA), PTC2 or IRE1 in

CC the cell, e.g. by gene overexpression. The cell from which the protein is

CC secreted can be any cell having an UPR, such as mammalian cells, insect

CC cells, yeast and filamentous fungi. The protein of interest can be any

CC secreted protein such as a therapeutic protein or an industrial enzyme,

CC e.g. lipase, cellulase, endoglucanase-H, protease, carboxylate, alpha-

CC reductase, oxidase, isomerase, transferase, kinase, phosphatase, alpha-

CC amylase, glucoamylase, lignocellulose hemicellulase, pectinase and

CC ligninase (claimed). (Updated on 11-SEP-2003 to standardise OS field)

XX

SQ Sequence 350 AA;

Query Match 99.4%; Score 1755.5; DB 4; Length 350;

Best Local Similarity 99.7%; Pred. No. 6.8e-139;

Matches 349; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 MKSADRFSPVKMEDAFANS-PTTPSLEVPVLTVPADTSLRTKNVVAQTKPEKKPAKCR 59

DB 1 MKSADRFSPVKMEDAFANSLEPTTPSLEVPVLTVPADTSLRTKNVVAQTKPEKKPAKCR 60

QY 60 KSWGQELVPKTNLPKRAKTEDEKEQRRIRVLRNRAAAQTSRKRKLEMEKLESEKI 119

DB 61 KSWGQELVPKTNLPKRAKTEDEKEQRRIRVLRNRAAAQTSRKRKLEMEKLESEKI 120

QY 120 DMEQONQFLQLRQAQEAENNRLSQQAQLSAEVRGSRHSTPTSSPASVSTLTPTLPK 179

DB 121 DMEQONQFLQLRQAQEAENNRLSQQAQLSAEVRGSRHSTPTSSPASVSTLTPTLPK 180

QY 180 QEGDEVPLDRIPFPPTPSVTDYPTLKPSLSAESPDLTQHPAVSVGLEGDESALTFLDLG 239

DB 181 QEGDEVPLDRIPFPPTPSVTDYPTLKPSLSAESPDLTQHPAVSVGLEGDESALTFLDLG 240

QY 240 ASIKHEPTHDLTAPLSDDDFRFLFNGDSSLESDDSLLEDGFAFDVLDGDLGSAFFPDSMV 299

DB 241 ASIKHEPTHDLTAPLSDDDFRFLFNGDSSLESDDSLLEDGFAFDVLDGDLGSAFFPDSMV 300

QY 300 DFDTEPVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDCQGIAAGSA 349

DB 301 DFDTEPVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDCQGIAAGSA 350

RESULT 5

AAE15372

ID AAE15372 standard; protein; 350 AA.

XX

AC AAE15372;

XX

XX 29-AUG-2003 (revised)

DT 07-MAR-2002 (first entry)

XX

DE Aspergillus nidulans hacA protein.

XX

XX Heterologous protein secretion; unfolded protein response; UPR; lipase;

KW cellulase; carbohydrase; industry; purification; hacA protein.

XX

OS Emericella nidulans.

XX

XX Key Location/Qualifiers

FT Binding-site 53..116

FT /label= DNA-binding_domain

XX

XX US2001034045-A1.

PN

XX 25-OCT-2001.

PD

XX

PF 23-MAR-2001; 2001US-00816277.

XX

PR 24-MAR-2000; 2000US-00534692.

XX

XX (GENV) GENENCOR INT INC.

PA

PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;

XX

XX WPI: 2002-033728/04.

DR N-PSDB; AAD24596.

XX

XX Increasing secretion of heterologous proteins e.g. lipase and cellulase

PT in eukaryotic cells useful in industry to increase production and

PT facilitate purification, by inducing an elevated unfolded protein

PT response.

XX

PS Claim 39; Fig 8; 56pp; English.

XX

CC The present invention relates to methods for increasing the secretion of

CC heterologous protein in eukaryotic cells by inducing an elevated unfolded

CC protein response (UPR). The method involves inducing the elevated UPR by

CC increasing the presence of proteins such as HAC1, HAC2, PTC2 or IRE1 in

CC cells. The method and sequences are useful for increasing the secretion

CC of heterologous proteins (e.g. lipase, cellulase, carboxylate) in

CC eukaryotic cells useful in industry to increase protein yields and to

CC facilitate purification. The present sequence is Aspergillus nidulans

CC hacA protein. (Updated on 29-AUG-2003 to standardise OS field)

XX

SQ Sequence 350 AA;

Query Match 99.4%; Score 1755.5; DB 5; Length 350;

Best Local Similarity 99.7%; Pred. No. 6.8e-139;

Matches 349; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 MKSADRFSPVKMEDAFANS-PTTPSLEVPVLTVPADTSLRTKNVVAQTKPEKKPAKCR 59

DB 1 MKSADRFSPVKMEDAFANSLEPTTPSLEVPVLTVPADTSLRTKNVVAQTKPEKKPAKCR 60

QY 60 KSWGQELVPKTNLPKRAKTEDEKEQRRIRVLRNRAAAQTSRKRKLEMEKLESEKI 119

DB 61 KSWGQELVPKTNLPKRAKTEDEKEQRRIRVLRNRAAAQTSRKRKLEMEKLESEKI 120

QY 120 DMEQONQFLQLRQAQEAENNRLSQQAQLSAEVRGSRHSTPTSSPASVSTLTPTLPK 179

DB 121 DMEQONQFLQLRQAQEAENNRLSQQAQLSAEVRGSRHSTPTSSPASVSTLTPTLPK 180

QY 180 QEGDEVPLDRIPFPPTPSVTDYPTLKPSLSAESPDLTQHPAVSVGLEGDESALTFLDLG 239

DB 181 QEGDEVPLDRIPFPPTPSVTDYPTLKPSLSAESPDLTQHPAVSVGLEGDESALTFLDLG 240

QY 240 ASIKHEPTHDLTAPLSDDDFRFLFNGDSSLESDDSLLEDGFAFDVLDGDLGSAFFPDSMV 299

DB 241 ASIKHEPTHDLTAPLSDDDFRFLFNGDSSLESDDSLLEDGFAFDVLDGDLGSAFFPDSMV 300

QY 300 DFDTEPVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDCQGIAAGSA 349

DB 301 DFDTEPVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDCQGIAAGSA 350

RESULT 6

AAAB2977

ID AAAB2977 standard; protein; 342 AA.

XX

XX AAAB2977;

XX

XX 11-SEP-2003 (revised)

DT 21-DEC-2001 (first entry)

XX

DE Aspergillus niger hacA, involved in unfolded protein response.

XX

KW HacA; transcription factor; unfolded protein response; protein secretion.

XX

OS Aspergillus awamori.

DE Trichoderma reesei HAC1 protein.
XX Heterologous protein secretion; unfolded protein response; UPR; lipase;
KW cellulase; carbohydrase; industry; purification; HAC1 protein.
KW Hypocrea jecorina.
OS
FH Key Location/Qualifiers
FT Binding-site 84..147
FT /label= DNA-binding_domain
XX
XX US2001034045-A1.
XX
XX 25-OCT-2001.
XX
XX 23-MAR-2001; 2001US-00816277.
XX
XX 24-MAR-2000; 2000US-00534692.
XX
XX (GEMV) GENENCOR INT INC.
XX
XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX WPI; 2002-033728/04.
XX N-PSDB; AAD24595.
DR
DR Increasing secretion of heterologous proteins e.g. lipase and cellulase
PT in eukaryotic cells useful in industry to increase production and
PT facilitate purification, by inducing an elevated unfolded protein
PT response.
XX
XX Claim 39; Fig 7; 56pp; English.
XX
XX The present invention relates to methods for increasing the secretion of
CC heterologous protein in eukaryotic cells by inducing an elevated UPR by
CC protein response (UPR). The method involves inducing the elevated UPR by
CC increasing the presence of proteins such as HAC1, HACA, PTC2 or IRE1 in
CC cells. The method and sequences are useful for increasing the secretion
CC of heterologous proteins (e.g. lipase, cellulase, carbohydrase) in
CC eukaryotic cells useful in industry to increase protein yields and to
CC facilitate purification. The present sequence is Trichoderma reesei HAC1
CC protein. (Updated on 29-AUG-2003 to standardise OS field)
XX
XX Sequence 451 AA;
XX
XX Query Match 24.6%; Score 434; DB 5; Length 451;
XX Best Local Similarity 32.4%; Pred No. 1.4e-27;
XX Matches 149; Conservative 50; Mismatches 109; Indels 152; Gaps 21;
QY 5 DRFSFQMEDAFANGSTPTPSLEVPVLTVPADTSRLTKNVVAQ----- 47
DB 26 DNFTSL-----FADS--TPS-----TLNPRD--NMTPDSVADIDSRLSVIPESQDAEDD 70
QY 48 -----TKP--BEKKPAKRKSWGOELPVKTNLPKRKATEDEKORRIEVLNRA 98
DB 71 ESHSTSATAPSTSEKPKVKRKSQGLPEPKTNLPKRKATEDEKORRVERVLNRR 130
QY 99 AAOQSRERKRLMEKLESEK-----IDMEQONQFLQRLAQMEANNRLS----- 143
DB 131 AAOQSRERKRLVEALEKRNKELETLINQKTNLILVEELNFRSSGVVTRSSPLDS 190
QY 144 -QQAQLSAEVRGSRH-----STPTSS-----SPASVSPTLTPTLFKQ---- 180
DB 191 LQDSITLSQQLFGSRDQGTMSNPEQSLMDQIMRSAANPTVNPASLSPLSPISDFQTK 250
QY 181 EGDEVPDLRIPTPTSVTDYSTLKPSSLAE-----SPDITQHPAVSVGLEGDESA 232
DB 251 EDEEQADE---DEEMEQTWHETKEAANAARKNKSKRSVSTSTORPAVSGG----DAA 303
QY 233 LTLF--DLGAS--IKHEPTDLTAPLSDDDFRLLFNGDSSLESPLLEDGFADFVLDSDG 289
DB 304 VPVFSDDAGANCLGLDPVHQDDGPFSS---IGHFGLSALDADRYLLE-----SQ 350

QY 290 LSAPFPDPMVDFO-----TEPVLE-DLEOTNGLSDSASCKAASL----- 328
DB 351 LLASFNASTVDDDDYLAGDSACFTNPLPSVDYDFDINFLTDANHAAYDIVAASNYAAAD 410
QY 329 -----QPSHGASTSRCDGGGIAAG 347
DB 411 RELDLEIHPENQIPSRHSIQQPQSGASSHGCGDDGGIAGV 450
RESULT 12
ADS12801
ID ADS12801 standard; protein; 451 AA.
XX
XX ADS12801;
XX
XX 16-DEC-2004 (first entry)
XX Trichoderma reesei hac1 chaperone and foldase.
XX
XX unfolded protein response; UPR; HAC1; PTC2;
KW unfolded protein response modulator; enzyme production; hac1; chaperone;
KW foldase; enzyme.
XX
XX Hypocrea jecorina.
XX
XX US2004186070-A1.
XX
XX 23-SEP-2004.
XX
XX 15-SEP-2003; 2003US-00663450.
XX
XX 24-MAR-2000; 2000US-00534692.
XX 23-MAR-2001; 2001US-00816277.
XX (GEMV) GENENCOR INT INC.
XX
XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
PI WPI; 2004-707924/69.
XX N-PSDB; ADS12800.
DR
DR Increasing the secretion of a heterologous protein, such as a therapeutic
PT or an industrial enzyme, in genetically modified eukaryotic cells by
PT inducing an elevated unfolded protein response (UPR).
XX
XX Example 3; SEQ ID NO 2; 83pp; English.
XX
XX The invention describes a method of increasing the secretion of a
CC heterologous protein in a eukaryotic cell, comprising inducing an
CC elevated unfolded protein response (UPR). Also described are: an isolated
CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
CC unfolded protein response and has less than 50% similarity to yeast HAC1
CC protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
CC protein induces unfolded protein response and where the HAC1 protein
CC comprises a DNA binding region that has greater than 70% similarity to
CC the DNA binding region of filamentous fungi HAC1 protein; a protein
CC having unfolded protein response inducing activity and having greater
CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
CC respectively) as given in the specification; a protein having an amino
CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
CC response and has at least 70% similarity to a fully defined amino acid
CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
CC specification; an isolated nucleic acid encoding a IRE1 protein that
CC modulates unfolded protein response and has at least 60% similarity to a
CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
CC having unfolded protein response modulating activity and having greater
CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
CC heterologous nucleic acid encoding a protein having unfolded protein
CC response modulating activity and a heterologous nucleic acid encoding a
CC protein of interest to be secreted. The methods and compositions of
CC genetically manipulating cells to have an elevated unfolded protein

QY	329	-----QPSHGASTSRCDGQIAAG 347	
Db	411	RELLEIHPENQIPSRHSIQQPSGASHCCDDGGIAGV 450	
RESULT 14			
AAAB82975			
ID	AAAB82975	standard; protein, 450 AA.	
AC	AAAB82975;		
XX			
DT	11-SEP-2003	(revised)	
DT	21-DEC-2001	(first entry)	
XX			
DE	Trichoderma reesei HAC1, involved in unfolded protein response.		
XX			
KW	HAC1; transcription factor; unfolded protein response; protein secretion.		
XX			
OS	Hypocrea jecorina.		
XX			
FH	Key	Location/Qualifiers	
FT	Domain	84..147	
FT		/label= DNA binding domain	
XX			
PN	WO200172783-A2.		
XX			
PD	04-OCT-2001.		
XX			
PF	23-MAR-2001; 2001WO-US009401.		
XX			
PR	24-MAR-2000; 2000US-00534692.		
XX			
PA	(GENV) GENENCOR INT INC.		
XX			
PI	Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;		
XX			
DR	WPI; 2001-626252/72.		
DR	N-PSDB; AAH26931.		
XX			
PT	Increasing secretion of heterologous proteins e.g. lipase and cellulase		
PT	in eukaryotic cells useful in industry to increase production and		
PT	facilitate purification, by inducing an elevated unfolded protein		
PT	response.		
XX			
PS	Claim 54; Fig 7A-B; 89pp; English.		
XX			
CC	The present sequence is that of the HAC1 protein of Trichoderma reesei, as deduced from the newly isolated HAC1 gene (see AAH26931). HAC1 protein is a transcription factor involved in the unfolded protein response (UPR). The invention provides methods for increasing the secretion of a heterologous protein in a cell by inducing an elevated UPR. This can be achieved by modulating the activity of HAC1, PTC2 or IRE1 in the cell, e.g. by gene overexpression. The cell from which the protein is secreted can be any cell having an UPR, such as mammalian cells, insect cells, yeast and filamentous fungi. The protein of interest can be any secreted protein such as a therapeutic protein or an industrial enzyme, e.g. lipase, cellulase, endoglucanase-H, protease, carbohydratase, reductase, oxidase, isomerase, transferase, kinase, phosphatase, alpha-amylase, glucoamylase, lignocellulose hemicellulase, pectinase and ligninase (claimed). (Updated on 11-SEP-2003 to standardise OS field)		
XX			
SQ	Sequence 450 AA;		
Query Match			
Best Local Similarity		23.8%;	Score 419.5; DB 4; Length 450;
Matches 148; Conservative 50; Mismatches 109; Indels 153; Gaps 22;			
QY	5	DRFSPVKMEDANSPPTPSLEVPVLTVPADTSLRTKNWAQ-----	47
Db	26	DNFTSL-----FADS--TFP-----TLNPRD--WMTSPSVADIDSRLSVIPESQDAEDD 70	
QY	48	-----TKP--EEKPAKRRKSWGQELPVPKTNLPKRKRAKTEDEKEQRIERVLRNRA 98	

THIS PAGE BLANK (USPTO)

Result No.	Score	Query Match	Length	DB	ID	Description	
1	429.5	18.7	409	2	US-09-533-029-104	Sequence 104, App	
2	258.5	11.3	248	2	US-09-248-796A-18840	Sequence 18840, A	
3	149.5	6.5	901	2	US-09-248-796A-14747	Sequence 14747, A	
4	148.5	6.5	650	2	US-09-487-558B-430	Sequence 430, App	
5	147.5	6.4	2058	2	US-09-949-016-6835	Sequence 6835, Ap	
6	147.5	6.4	2111	2	US-09-949-016-10199	Sequence 10199, A	
7	146.5	6.4	667	2	US-09-248-796A-14492	Sequence 14492, A	
8	143.5	6.3	671	2	US-10-012-231A-308	Sequence 308, App	
9	143.5	6.3	671	2	US-10-015-389A-308	Sequence 308, App	
10	143.5	6.3	671	2	US-10-006-768A-308	Sequence 308, App	
11	143.5	6.3	671	2	US-10-015-671A-308	Sequence 308, App	
12	143.5	6.3	671	2	US-10-015-393A-308	Sequence 308, App	
13	143.5	6.3	671	2	US-10-011-833A-308	Sequence 308, App	
14	143.5	6.3	671	2	US-10-006-041A-308	Sequence 308, App	
15	143.5	6.3	671	2	US-10-012-064A-308	Sequence 308, App	
16	141	6.1	778	2	US-09-248-796A-16358	Sequence 16358, A	
17	140.5	6.1	529	2	US-09-248-796A-16357	Sequence 16357, A	
18	138.5	6.0	2137	2	US-09-134-001C-4463	Sequence 4463, Ap	
19	138	6.0	619	2	US-09-066-046-2	Sequence 2, Appli	
20	136.5	6.0	605	2	US-09-394-645-2	Sequence 2, Appli	
21	136.5	6.0	605	2	US-09-243-560B-2	Sequence 2, Appli	
22	136.5	6.0	605	2	US-09-832-129-37	Sequence 37, Appli	
23	136	5.9	352	2	US-09-248-796A-18019	Sequence 18019, A	
24	136	5.9	521	1	US-08-021-684C-2	Sequence 2, Appli	
25	136	5.9	521	1	US-09-005-970-2	Sequence 2, Appli	
26	136	5.9	521	2	US-09-407-715-2	Sequence 2, Appli	
27	135.5	5.9	1878	2	US-09-949-016-8902	Sequence 8902, Ap	

```
QY 177 SSGVTRSSPLDLSQDS-----ITLSQQLFGSRDQTMNPEQSLMDQIMRSAANPTVNP 232
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 169 TNFLSYSDSTPDISDSQSLPSTFKQLFNAQD--ELCRP---ISPQSIGPLTSRTVDP 223
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 233 ASLSPSLPISDKFOTKEDEBEQADEMEQOTWHTKEAAAAKSKQSRVSTDSIQ 292
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 224 STLSP-----KSLSPDSSNSN-----SSDMTQ 246
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 293 RPAV 296
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 247 HPAV 250
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 2
US-09-248-796A-18840
; Sequence 18840, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18840
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Candida albicans
; NAME/KEY: UNSURE
; LOCATION: (221)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unkno
US-09-248-796A-18840
Query Match 11.3%; Score 258.5; DB 2; Length 248;
Best Local Similarity 29.7%; Pred. No. 4.6e-14;
Matches 80; Conservative 43; Mismatches 65; Indels 81; Gaps 10;

QY 31 LEADSTPSTLNPRDM--MTPDSVADIDSRSLVSPESQDAEDDESHSTSATAPSTSEKPV 88
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 21 LTVDNNTNTSNDLDSVAFTSLM-----TSITTSFSMS----- 54
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 89 KKRKSGQVLP-EP---KTNLPPRKAKTEDEKQRRVERVLNRRAAQSSRRKRLEVE 144
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 55 TSTSSHNTLIDDPATFKSTLPKRKAQTQEEKQKRIILNRRAAHASREKKRKHYE 114
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 145 ALEKRNKELETLINQKTNLILVEELNFRSSGVVTRSS-----SPLDSIQDSITLSQ 199
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 115 YLENYVLKLET---NUMKLN-----NNYNOAFELLTKDNQELLGSKLEVLDVSDLKE 164
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 200 QLFGSRDQTMNPEQSLMDQIMRSAANPTVNPASLSPISDKFOTKEDEBEQADE 259
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 165 QIHSNWSGTRSHNKS-----NDEDIEEDE 191
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 260 DEMEQOTWHTKEAAAAKSKQSRVST 288
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 192 DEHLEB-GHVEQEIKEEPVSKRKLTNT 219
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 3
US-09-248-796A-14747
; Sequence 14747, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
```

```
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14747
; LENGTH: 901
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-14747
Query Match 6.5%; Score 149.5; DB 2; Length 901;
Best Local Similarity 21.7%; Pred. No. 0.00061;
Matches 78; Conservative 59; Mismatches 142; Indels 81; Gaps 15;

QY 30 SLFADSTPSTLNPRDMTPDSVADIDSRSLVSPESQDAEDDESH-----STSATA 79
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 455 ALISDDPYSTIEQLDITMVEVLHIGDNTLTDTKSDTKDSEKKEKFAEPKSVDTKTELE 514
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 80 PSTSEKPVKKKSGWG-QVLPEPKT--NLPPRKRAKTED-----EKEQRRVERVLNRN- 128
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 515 TKTSEKS--EESKELGVKLEKVDTEKEIPVEESKKVDDNDDETDLKESKETKVEKASDKP 572
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 129 -----RRAAQSSRRERKRLEVEALEKRNKELETLINNVQKTNLILVEELNFRFRS 177
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 573 ITHESLIDEDIEKAKHQSNKDDANDTVLAE-----EGALVNEDDYDLVDKKELEK- 623
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 178 SGVTRSSPLD--SLQDSITLSQQLFGSRDQTMNPEQSLMDQIMRSAANPTVNPASL 235
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 624 EGITTEEAPLNKEKKEKTKQVEDVLESSVKKDEKPILDVESKEAKKAADDEEKAASE 683
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 236 SPLPPI--SKKEFOTKEDE-----EQADED-----ERMEQOTWH----- 268
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 684 QPPLDIVKSSKETEKETDETTKSELPGTKVEDSEKDTTLTKSEVESELEKSEBQPLDI 743
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 269 -----ETKEAAAAKSKQSRVS--TDSTORPAVSIIGDAAVPVFSDDAGANCLGLDP 320
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 744 KKEVVEYTKDDVATEKSKDVEQAVSVSTYKTTKPEV-LETEKPKPAVADDDDLDDDISP 802
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
US-09-487-558B-430
; Sequence 430, Application US/09487558B
; Patent No. 6949356
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 109272.130
; CURRENT APPLICATION NUMBER: US/09/487,558B
; CURRENT FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/487,558
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 446
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 430
; LENGTH: 650
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-487-558B-430
```

Query Match 6.5%; Score 148.5; DB 2; Length 650;
Best Local Similarity 17.6%; Pred.No.0.00046;
Matches 106; Conservative 85; Mismatches 171; Indels 239; Gaps 23;

QY 44 DMWTPDSVADTDSRLSVIPESODAEDESHSSTAPSTSEKKPVKPKRKSOGVLPEPKT 103
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 11 DVVSFGSLAEPG-----SKSRHDIENEHRTGTDRGEDSDQPCKKGS----- 54
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 104 NLPPRKRAKTEDEKEORRVVERVLNRRAAQSSRRKERLEVALEAKRNKELETLLINVKQT 163
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 55 -----KTSKGQQLDPETKQKRTAQNRAQAQRPRERKRWKELEKVKVQSL-----SIQQQ 105
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 164 N-----LILAVEELNRF-----RSSGVTVTRSSPLDS 190
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 106 NEVEATFLRDQILITLVNELKLYRPETRNDISKVLEYLARDPNLHFHSKNVNHSNSEPIDT 165
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 191 LQDSI-----TLSQLFGSGDGTMSNPQS-----MDQIMRSAAN 227
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 166 PNDDIQENVKQKMFTFYPLDNNDNDNSKVVGKQLSPNDPSHSAHPINQTKKLSD 225
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 228 PTVPASLSPLPSID-----KEFQTKEBDEBQA 257
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 226 ATDSSSATLDSLNSNDVLNNTPNSSTSMDWLNDNVYTNRFSVGDGSGNSTKNLDSNMF 285
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 258 DEDEMEQTWE-----TXE-----AAAAEKNSKQSRVSTDSTQSPA 295
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 286 SDFNFENQFDEQVFSEPCSKNQNVCGTCQPIPKPIKPIKPIKPIKPIKPIKPIKPIKPI 342
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 296 VSIIGDAAPVFSFDAGANCLGLDPVHQDGPFSIGHGFG-LSAALDADRYYLESQLLAS 354
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 343 LTNTWESHNI-TDNTPANVIATDATKYENS-FS--GFGRGLGPDMSANHYVVNDNSTGS 397
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 355 NA-----STVDVDDL-----AGBSAACTNP----- 376
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 398 TDSTGSTGNKNKNKNNSDVLFPFISEPFDMNQVTNFFSPGSTGIGNNAAGTNPSLIQ 457
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 377 -----LPSD-----YDFDI-----NDFLTLD 391
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 458 SKKEDIPIFNANLAPPDNDNSTNIQLQPFSEBSOSQNKFDYDMFFRDSKSGNNLFGFEFLD 517
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 392 DANHAAYDIVAASNYAADREL---DLBIHOPE-----NQTPSRHSIQQPOSQASSHG 442
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 518 DD-----DDKKAANSDDESSLIKLNLINEPELPKQYLQSVPGNESEBISQKNGSSLQNA 572
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 443 D 443
|
DB 573 D 573

RESULT 5
US-09-949-016-6835
Sequence 6835, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CLO001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6835
LENGTH: 2058
TYPE: PRT
ORGANISM: Human

```

US-09-949-016-6835

Query Match          6.4%; Score 147.5; DB 2; Length 2058;
Best Local Similarity 23.5%; Pred. No. 0.003;
Matches 86; Conservative 58; Mismatches 119; Indels 103; Gaps 18;

Qy      91  RKSQGVLPPEKTNLPK-----RAKTEDEKEQRRVERVLNRRAAQSSRRKRLEVEA 145
Db      805  RRVYRQLLAEKREQEEKKQEEBKKREBERERERERAEALRAQOEETRKQOELEA 864
Qy      146  LEKRNKELE-TLLINVQKTNLLIVELNPRFRSSGVVTRSSPLDSLQSDITLSQQLFGS 204
Db      865  LQKSQKEAELTRELEKQENK-QVEEILREK-----EIEDLQRMKEQ-----906
Qy      205  RDGQTMNPEQSLMDQIMRSAAANTPVNPASLSPSPISPDKFQTKEEDEQA-----257
Db      907  ---QELSLTEASLQ-----KLERRDQELRLREEACRAAQEFLES 944
Qy      258  ---DEDEE-----MEQWTHETKEAAAAKEKNSQSRVSTSTQ---RPVYSIGGDAAPVVF 307
Db      945  LNFDEIDECVRNIERSLSVGSSEFSELAESACEKFNFPQYPPEEVDGFEADDDAF 1004
Qy      308  SDDAGANCLGDPHQDDGPFSIGHS-----FGLSAAALDA---DRYLLESQALLSPNA-ST 359
Db      1005  KD-----SPNPSEHGHSQDQRTSGTISDSDSEEDPYMNDTVVTPSPSADST 1050
Qy      360  V-----DDLYL-----AGDSAAACF---TNPLPS---DYDFDINDFLTDDANHAAYDIVA 402
Db      1051  VLLAPSVQDSGLNHSNGSGESTYCMQPQAGDLFSPDGDYDYDQDDY-EDGAIISGSSVTF 1109
Qy      403  ASNYAA 408
Db      1110  SNSYGS 1115

RESULT 6
US-09-949-016-10199
; Sequence 10199, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THERE
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10199
; LENGTH: 2111
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10199

Query Match          6.4%; Score 147.5; DB 2; Length 2111;
Best Local Similarity 23.5%; Pred. No. 0.0031;
Matches 86; Conservative 58; Mismatches 119; Indels 103; Gaps 18;

Qy      91  RKSQGVLPPEKTNLPK-----RAKTEDEKEQRRVERVLNRRAAQSSRRKRLEVEA 145
Db      858  RRVYRQLLAEKREQEEKKQEEBKKREBERERERERAEALRAQOEETRKQOELEA 917
Qy      146  LEKRNKELE-TLLINVQKTNLLIVELNPRFRSSGVVTRSSPLDSLQSDITLSQQLFGS 204
Db      918  LQKSQKEAELTRELEKQENK-QVEEILREK-----EIEDLQRMKEQ-----959
Qy      205  RDGQTMNPEQSLMDQIMRSAAANTPVNPASLSPSPISPDKFQTKEEDEQA-----257

```

```

US-09-949-016-6835

Query Match          6.4%; Score 147.5; DB 2; Length 2058;
Best Local Similarity 23.5%; Pred. No. 0.003;
Matches 86; Conservative 58; Mismatches 119; Indels 103; Gaps 18;

Qy      91  RKSQGVLPPEKTNLPK-----RAKTEDEKEQRRVERVLNRRAAQSSRRKRLEVEA 145
Db      805  RRVYRQLLAEKREQEEKKQEEBKKREBERERERERAEALRAQOEETRKQOELEA 864
Qy      146  LEKRNKELE-TLLINVQKTNLLIVELNPRFRSSGVVTRSSPLDSLQSDITLSQQLFGS 204
Db      865  LQKSQKEAELTRELEKQENK-QVEEILREK-----EIEDLQRMKEQ-----906
Qy      205  RDGQTMNPEQSLMDQIMRSAAANTPVNPASLSPSPISPDKFQTKEEDEQA-----257
Db      907  ---QELSLTEASLQ-----KLERRDQELRLREEACRAAQEFLES 944
Qy      258  ---DEDEE-----MEQWTHETKEAAAAKEKNSQSRVSTSTQ---RPVYSIGGDAAPVVF 307
Db      945  LNFDEIDECVRNIERSLSVGSSEFSELAESACEKFNFPQYPPEEVDGFEADDDAF 1004
Qy      308  SDDAGANCLGDPHQDDGPFSIGHS-----FGLSAAALDA---DRYLLESQALLSPNA-ST 359
Db      1005  KD-----SPNPSEHGHSQDQRTSGTISDSDSEEDPYMNDTVVTPSPSADST 1050
Qy      360  V-----DDLYL-----AGDSAAACF---TNPLPS---DYDFDINDFLTDDANHAAYDIVA 402
Db      1051  VLLAPSVQDSGLNHSNGSGESTYCMQPQAGDLFSPDGDYDYDQDDY-EDGAIISGSSVTF 1109
Qy      403  ASNYAA 408
Db      1110  SNSYGS 1115

RESULT 6
US-09-949-016-10199
; Sequence 10199, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THERE
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10199
; LENGTH: 2111
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10199

Query Match          6.4%; Score 147.5; DB 2; Length 2111;
Best Local Similarity 23.5%; Pred. No. 0.0031;
Matches 86; Conservative 58; Mismatches 119; Indels 103; Gaps 18;

Qy      91  RKSQGVLPPEKTNLPK-----RAKTEDEKEQRRVERVLNRRAAQSSRRKRLEVEA 145
Db      858  RRVYRQLLAEKREQEEKKQEEBKKREBERERERERAEALRAQOEETRKQOELEA 917
Qy      146  LEKRNKELE-TLLINVQKTNLLIVELNPRFRSSGVVTRSSPLDSLQSDITLSQQLFGS 204
Db      918  LQKSQKEAELTRELEKQENK-QVEEILREK-----EIEDLQRMKEQ-----959
Qy      205  RDGQTMNPEQSLMDQIMRSAAANTPVNPASLSPSPISPDKFQTKEEDEQA-----257

```

Db 960 ---QELSLTEASIQ-----KLQERRDQBLRLREEEACRAAQEFLES 997
QY 258 ---DEDEE---MEQTHWETKEAAAKENKSKQSRVSTDSQ---RPAVSIIGDAAVPVF 307
Db 998 LNFDEIDECVRIERSLSVGSFELSAAECBEPNFNFQPYPEEVEDEGEFADDDAF 1057
QY 308 SDDAGANCLGLDPVHQDDGPPFSGHGS-----FGLSAAALDA---DRYLLESQILLASPNA-ST 359
Db 1058 KD-----SPNPSEHGSHDQRTSGIRTSDDSEEDPYWMDIVVTSPSAUST 1103
QY 360 V-----DDDYL-----AGDSAACF---TNPLPS---DYDFDINDFLTDANHAAYDIVA 402
Db 1104 VLLAPSVQDSGLHNSSGESTYCMQPNAGDLPSPDGDYDYDQDDY-EDGALTSGSVTF 1162
QY 403 ASNYAA 408
Db 1163 SNSYGS 1168
RESULT 7
US-09-248-796A-14492
; Sequence 14492, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248, 796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1998-02-13
; PRIOR FILING DATE: 1998-02-13
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14492
; LENGTH: 667
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-14492
Query Match 6.4%; Score 146.5; DB 2; Length 667;
Best Local Similarity 23.6%; Pred. No. 0.00071;
Matches 87; Conservative 62; Mismatches 146; Indels 73; Gaps 16;
QY 56 SRLSVIPESQDAEDDESHSTATAPSTSEK-KPVKKRKSQGVLPPEKTNLPPEK----- 109
Db 139 SAVTASSATPEPENESLAASEQNEAPKAPKAKKPKAGLAAALKKQLELKKQLEEE 198
QY 110 RAKTEDEKEORV-----ERVLNRRAAQSSRRERKLEVEALE-----KRKNELET 155
Db 199 QRRLEEEERLEERLAEEBAEAKKAAAKKRRERLUKKQLEKQLEKQLEK 258
QY 156 L-----LI-----NVQKTNLILVEELNRRPRSSGVVT--RSSPSLDSLQDSITLSQQLRG 203
Db 259 LQERRRQQLLAGNVSVAGLSKTEGDAPKPKVYVTKKSKTKPTFIQKTPV-SKAPVK 317
QY 204 SRDQTMNPEOSLMDQIMRSANPTVNPDA-SLSPLPPIPSKPEFQTKKEDEEQADEEBE 262
Db 318 KDDGE-----DEALVDDMEKMLDAPVADPVADEAALADEGDEEVEGEAEDEEVEGEDEID 372
QY 263 -----MEQTHWETKEAAA--AKEKSKQSRVSTDSQTPAVSISGGDAAVPVPSD 309
Db 373 TIVEDEQEKKEAEEAAKAKESARKAKKEAAKQEEBSKQLAAQK---KAAAPSEKD 429
QY 310 DAGANCLGLDPVHQDDGPPFSI-----GHSFGLSAAALDA-----DRYLLESQILLA 353
Db 430 LRSPICCILG--HVDGTGKLLDKIRQTNVQGEAGGITQIQIGATVPVDAIKQKTAVMA 487
QY 354 SPNASTVD 361
Db 488 KYEQKTFD 495

RESULT 8
US-10-012-231A-308
; Sequence 308, Application US/10012231A
; Patent No. 6924355
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830PIC23
; CURRENT APPLICATION NUMBER: US/10/012, 231A
; CURRENT FILING DATE: 2002-06-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 308
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-231A-308
Query Match 6.3%; Score 143.5; DB 2; Length 671;
Best Local Similarity 24.8%; Pred. No. 0.0013;
Matches 79; Conservative 56; Mismatches 138; Indels 45; Gaps 15;
QY 2 AFQSSPLVKPEASPAESFLS-APGDNFTSLPADSTP-----STLNPRDMWTPD 49
Db 170 ARKASSDLDQASVSPSEENSESSESEKTSQDPTPEKAAVAPRRGFLGRKKKKAP 229
QY 50 SVADIDSRL---SVIPESQDAEDDESHSTATAPSTSE---KKPVKKRKSQGVLPPEKPT 103
Db 230 SASDSKSDKADSGAKPEPVAMARSASSSSSSSSSDSDSVSVKPPRGRKPAEKPLPKPRG 289
QY 104 NLPPKR--AKTEDEKEORRVERVLNRRRAAQSSRRERKLEVEALEKRNKELETLLINVQ 161
Db 290 RKPKEPFPSSSSSDSDSDVDRISEWKR-----RDEARRRRELEARRRREOEELRLREQ 345
QY 162 KTNLILVEELNRRFRSSGVVTRSS--SPLDSL-QDSITLSQQLFGSRDQ-----TWS 211
Db 346 EKE---EKERRRRADRGEARGSGSGGDELREDDPEVKKR---GRKGRGPPSSSSS 399
QY 212 NPEQSLMDQIMRSANPTVNPASLPSLPISDKKEFQTKKEDEEQADEEBEOTWHTK 271
Db 400 EPEAELEREAKKSACKP--QSSSTEPARKP-GQEKERVPEEKQQA-KPVKVERT-RKGS 454
QY 272 EAAAKEKNSQSRVSTD 289
Db 455 EGFSMDRKVEKKKEPSVE 472
RESULT 9
US-10-015-389A-308
; Sequence 308, Application US/10015389A
; Patent No. 6936436
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone

```
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830PIC48
; CURRENT APPLICATION NUMBER: US/10/015,389A
; CURRENT FILING DATE: 2002-06-25
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 308
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-015-389A-308

Query Match      6.3%; Score 143.5; DB 2; Length 671;
Best Local Similarity 24.8%; Pred. No. 0.0013;
Matches 79; Conservative 56; Mismatches 138; Indels 45; Gaps 15;

QY 2 AFQSSPLVKFEASPAESFLS-APGDNFTSLPADSTP-----STLNPRDMMTDP 49
Db 170 ARKASSDLQASVSPSEENSESSSEKTSQDFTPEKKAAVRAPRRGPLGRKKKAP 229
QY 50 SVADIDSRL---SVIPESQDAEDDHSHTSATPTSE---KKPVKKRSWQVLPKPT 103
Db 230 SASDSKADSGAKPEPVAMARSASSSSSSSSSDSDSVSKKPPRGRKPAEKPLPKPRG 289
QY 104 NLPPRKR--AKTEDEKEQRRVRLNRRAAQSSRRERKLEVEALEKRNKELETLLINQV 161
Db 290 RKPKPERPPSSSSSDSDSDVDRISEWKR-----RDARRRELEARRRREBEELRLREQ 345
QY 162 KTNLILVELNFRSSGVVTRSS--SPLDSL-QDSITLSQOLFGRDQ-----TWS 211
Db 346 EKE---EKERRERADRGEAERGSGSGDELREDEPVKKR---GRKGRGPPSSSDS 399
QY 212 NPEQSLMQIWRSAANPTVNPASLSPSPPIIDKFEQYKEDEEEOADEEEMQTHETK 271
Db 400 EPEALELEAKKSARKP--QSSSTEPARKP-GQKEKRVPRPEKQQA-KPVKVERT-RKRS 454
QY 272 EAAAAKEKNSQSRVSTD 289
Db 455 EGFSMDRKVEKKKEPSVE 472

RESULT 10
US-10-005-768A-308
; Sequence 308, Application US/10006768A
; Patent No. 6936597
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830PIC47
; CURRENT APPLICATION NUMBER: US/10/015,671A
; CURRENT FILING DATE: 2001-12-11
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 308
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-015-671A-308

Query Match      6.3%; Score 143.5; DB 2; Length 671;
Best Local Similarity 24.8%; Pred. No. 0.0013;
```

```
; CURRENT APPLICATION NUMBER: US/10/006,768A
; CURRENT FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 477
; Prior Application removed - See File Wrapper or Palm
; SEQ ID NO 308
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-006-768A-308

Query Match      6.3%; Score 143.5; DB 2; Length 671;
Best Local Similarity 24.8%; Pred. No. 0.0013;
Matches 79; Conservative 56; Mismatches 138; Indels 45; Gaps 15;

QY 2 AFQSSPLVKFEASPAESFLS-APGDNFTSLPADSTP-----STLNPRDMMTDP 49
Db 170 ARKASSDLQASVSPSEENSESSSEKTSQDFTPEKKAAVRAPRRGPLGRKKKAP 229
QY 50 SVADIDSRL---SVIPESQDAEDDHSHTSATPTSE---KKPVKKRSWQVLPKPT 103
Db 230 SASDSKADSGAKPEPVAMARSASSSSSSSSSDSDSVSKKPPRGRKPAEKPLPKPRG 289
QY 104 NLPPRKR--AKTEDEKEQRRVRLNRRAAQSSRRERKLEVEALEKRNKELETLLINQV 161
Db 290 RKPKPERPPSSSSSDSDSDVDRISEWKR-----RDARRRELEARRRREBEELRLREQ 345
QY 162 KTNLILVELNFRSSGVVTRSS--SPLDSL-QDSITLSQOLFGRDQ-----TWS 211
Db 346 EKE---EKERRERADRGEAERGSGSGDELREDEPVKKR---GRKGRGPPSSSDS 399
QY 212 NPEQSLMQIWRSAANPTVNPASLSPSPPIIDKFEQYKEDEEEOADEEEMQTHETK 271
Db 400 EPEALELEAKKSARKP--QSSSTEPARKP-GQKEKRVPRPEKQQA-KPVKVERT-RKRS 454
QY 272 EAAAAKEKNSQSRVSTD 289
Db 455 EGFSMDRKVEKKKEPSVE 472

RESULT 11
US-10-015-671A-308
; Sequence 308, Application US/10015671A
; Patent No. 6946263
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830PIC47
; CURRENT APPLICATION NUMBER: US/10/015,671A
; CURRENT FILING DATE: 2001-12-11
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 308
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-015-671A-308

Query Match      6.3%; Score 143.5; DB 2; Length 671;
Best Local Similarity 24.8%; Pred. No. 0.0013;
```

Matches 79; Conservative 56; Mismatches 138; Indels 45; Gaps 15;
QY 2 AFQSSPLVKEASPAESFLS-APGDNFTSLFADSTP-----STLNPRDMWTPD 49
Db 170 ARKASSDLQASVSPSEENSESESEKTSQDQFTPEKKAAPRAVRPRGPGIGRKKKAP 229
QY 50 SVADIDSRL---SVIPESQDAEDDESHSTSATAPSTSE---KKPVKKRKSQGVLPPEPKT 103
Db 230 SASDSKADSDGAKPEPVMARSASSSSSSSSSDSDSVVKPPRGRKPAEKPLPKPRG 289
QY 104 NLPPRKR--AKTEDEKEQRRVERVLNRRAAQSRRERKLEVEALEKRNKELETLLINQV 161
Db 290 RKPKPERPPSSSSSDSDSDVDRISEWKR---RDEARRRELEARRRQEEELRLREQ 345
QY 162 KTNLILVEELNFRSSGGVTRSS--SPLDSL-QDSITLSQQLFGSRDQ-----TMS 211
Db 346 EKE---EKERRRERADRGEARGSGGSGDELREDDPEVKR---GRKGRGPPSSSDS 399
QY 212 NPEQSLMQIMRSAAANTVNPASLSPSPPIISDKFQTKKEDEQADEDEMEQTHWHTK 271
Db 400 EPEAELEREAKSAKKP--QSSSTEPARKP-GQKEKRVPEEKQQA-KPVKVERT-RKRS 454
QY 272 EAAAAKEKNKQSRVSTD 289
Db 455 EGFSMDRKVKKKEPSVE 472

RESULT 12
US-10-015-393A-308
; Sequence 308, Application US/10015393A
; Patent No. 6951737
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Goddard, Audrey
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830P1C46
; CURRENT FILING DATE: 2002-06-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 308
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-015-393A-308
Query Match 6.3%; Score 143.5; DB 2; Length 671;
Best Local Similarity 24.8%; Pred. No. 0.0013;
Matches 79; Conservative 56; Mismatches 138; Indels 45; Gaps 15;
QY 2 AFQSSPLVKEASPAESFLS-APGDNFTSLFADSTP-----STLNPRDMWTPD 49
Db 170 ARKASSDLQASVSPSEENSESESEKTSQDQFTPEKKAAPRAVRPRGPGIGRKKKAP 229
QY 50 SVADIDSRL---SVIPESQDAEDDESHSTSATAPSTSE---KKPVKKRKSQGVLPPEPKT 103
Db 230 SASDSKADSDGAKPEPVMARSASSSSSSSSSDSDSVVKPPRGRKPAEKPLPKPRG 289
QY 104 NLPPRKR--AKTEDEKEQRRVERVLNRRAAQSRRERKLEVEALEKRNKELETLLINQV 161

Db 290 RKPKPERPPSSSSSDSDSDVDRISEWKR---RDEARRRELEARRRQEEELRLREQ 345
QY 162 KTNLILVEELNFRSSGGVTRSS--SPLDSL-QDSITLSQQLFGSRDQ-----TMS 211
Db 346 EKE---EKERRRERADRGEARGSGGSGDELREDDPEVKR---GRKGRGPPSSSDS 399
QY 212 NPEQSLMQIMRSAAANTVNPASLSPSPPIISDKFQTKKEDEQADEDEMEQTHWHTK 271
Db 400 EPEAELEREAKSAKKP--QSSSTEPARKP-GQKEKRVPEEKQQA-KPVKVERT-RKRS 454
QY 272 EAAAAKEKNKQSRVSTD 289
Db 455 EGFSMDRKVKKKEPSVE 472

RESULT 13
US-10-011-833A-308
; Sequence 308, Application US/10011833A
; Patent No. 6951920
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830P1C22
; CURRENT FILING DATE: 2002-06-25
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 308
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-011-833A-308
Query Match 6.3%; Score 143.5; DB 2; Length 671;
Best Local Similarity 24.8%; Pred. No. 0.0013;
Matches 79; Conservative 56; Mismatches 138; Indels 45; Gaps 15;
QY 2 AFQSSPLVKEASPAESFLS-APGDNFTSLFADSTP-----STLNPRDMWTPD 49
Db 170 ARKASSDLQASVSPSEENSESESEKTSQDQFTPEKKAAPRAVRPRGPGIGRKKKAP 229
QY 50 SVADIDSRL---SVIPESQDAEDDESHSTSATAPSTSE---KKPVKKRKSQGVLPPEPKT 103
Db 230 SASDSKADSDGAKPEPVMARSASSSSSSSSSDSDSVVKPPRGRKPAEKPLPKPRG 289
QY 104 NLPPRKR--AKTEDEKEQRRVERVLNRRAAQSRRERKLEVEALEKRNKELETLLINQV 161
Db 290 RKPKPERPPSSSSSDSDSDVDRISEWKR---RDEARRRELEARRRQEEELRLREQ 345
QY 162 KTNLILVEELNFRSSGGVTRSS--SPLDSL-QDSITLSQQLFGSRDQ-----TMS 211
Db 346 EKE---EKERRRERADRGEARGSGGSGDELREDDPEVKR---GRKGRGPPSSSDS 399
QY 212 NPEQSLMQIMRSAAANTVNPASLSPSPPIISDKFQTKKEDEQADEDEMEQTHWHTK 271
Db 400 EPEAELEREAKSAKKP--QSSSTEPARKP-GQKEKRVPEEKQQA-KPVKVERT-RKRS 454
QY 272 EAAAAKEKNKQSRVSTD 289

```
Db 455 EGFSMDRKVEKKEPSVE 472

RESULT 14
US-10-006-041A-308
; Sequence 308, Application US/10006041A
; Patent No. 6951921
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830PIC8
; CURRENT APPLICATION NUMBER: US/10/006,041A
; Prior Filing Date: 2001-12-06
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 308
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-006-041A-308

Query Match 6.3%; Score 143.5; DB 2; Length 671;
Best Local Similarity 24.8%; Pred. No. 0.0013;
Matches 79; Conservative 56; Mismatches 138; Indels 45; Gaps 15;

QY 2 AFQSSPLVKFEASPSEFLS-APGDNFTSLPADSTP-----STLNPRDMWTPD 49
Db 170 ARKASSDLQASVSPSEENSESESEKTSQDFTPEKKAAPRAVRPGLGGRKKKAP 229

QY 50 SVADIDSRLL---SVIPESQDAEDDESHSTSATAPSTSE---KKPVKKRKSQVLPPEPT 103
Db 230 SASDSKADSGAKPEPVAMARSASSSSSSSSSDSDSVSKPPGRKPAEKPLPKPRG 289

QY 104 NLPPRKR--AKTEDEKQRRVRLNRRAAQSSRRKRLEVEALEKKNKELETLLINQV 161
Db 290 RKPKPERPPSSSSSDSDSDVDRISEWKR-----RDEARRRLEARRRQEELRLREQ 345

QY 162 KTNLILVEELNRRSSSGVVTSS--SPLDSL-ODSITLSQQLFGSRDQ-----TMS 211
Db 346 EKE---EKERRRERADRGEAERGSGGSDLELREDDPEVKR---GRKGRGPPSSSDS 399

QY 212 NPEQSLMDQIMRSAANTVPNAPSLPPTSDKFEQTKEDDEQADDEMEQTHETK 271
Db 400 EPEALEEREAKKSARKP--QSSSTEPARKP-GQKEKRVPEKQQA-KPVKVERT-RKRS 454

QY 272 EAAAKEKNKSKQSRVSTD 289
Db 455 EGFSMDRKVEKKEPSVE 472

RESULT 15
US-10-012-064A-308
; Sequence 308, Application US/10012064A
; Patent No. 6953841
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830PIC19
; CURRENT APPLICATION NUMBER: US/10/012,064A
; Prior Filing Date: 2002-07-15
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 308
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-064A-308

Query Match 6.3%; Score 143.5; DB 2; Length 671;
Best Local Similarity 24.8%; Pred. No. 0.0013;
Matches 79; Conservative 56; Mismatches 138; Indels 45; Gaps 15;

QY 2 AFQSSPLVKFEASPSEFLS-APGDNFTSLPADSTP-----STLNPRDMWTPD 49
Db 170 ARKASSDLQASVSPSEENSESESEKTSQDFTPEKKAAPRAVRPGLGGRKKKAP 229

QY 50 SVADIDSRLL---SVIPESQDAEDDESHSTSATAPSTSE---KKPVKKRKSQVLPPEPT 103
Db 230 SASDSKADSGAKPEPVAMARSASSSSSSSSSDSDSVSKPPGRKPAEKPLPKPRG 289

QY 104 NLPPRKR--AKTEDEKQRRVRLNRRAAQSSRRKRLEVEALEKKNKELETLLINQV 161
Db 290 RKPKPERPPSSSSSDSDSDVDRISEWKR-----RDEARRRLEARRRQEELRLREQ 345

QY 162 KTNLILVEELNRRSSSGVVTSS--SPLDSL-ODSITLSQQLFGSRDQ-----TMS 211
Db 346 EKE---EKERRRERADRGEAERGSGGSDLELREDDPEVKR---GRKGRGPPSSSDS 399

QY 212 NPEQSLMDQIMRSAANTVPNAPSLPPTSDKFEQTKEDDEQADDEMEQTHETK 271
Db 400 EPEALEEREAKKSARKP--QSSSTEPARKP-GQKEKRVPEKQQA-KPVKVERT-RKRS 454

QY 272 EAAAKEKNKSKQSRVSTD 289
Db 455 EGFSMDRKVEKKEPSVE 472

Search completed: November 23, 2005, 03:27:32
```

```
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830PIC19
; CURRENT APPLICATION NUMBER: US/10/012,064A
; Prior Filing Date: 2002-07-15
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 308
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-064A-308

Query Match 6.3%; Score 143.5; DB 2; Length 671;
Best Local Similarity 24.8%; Pred. No. 0.0013;
Matches 79; Conservative 56; Mismatches 138; Indels 45; Gaps 15;

QY 2 AFQSSPLVKFEASPSEFLS-APGDNFTSLPADSTP-----STLNPRDMWTPD 49
Db 170 ARKASSDLQASVSPSEENSESESEKTSQDFTPEKKAAPRAVRPGLGGRKKKAP 229

QY 50 SVADIDSRLL---SVIPESQDAEDDESHSTSATAPSTSE---KKPVKKRKSQVLPPEPT 103
Db 230 SASDSKADSGAKPEPVAMARSASSSSSSSSSDSDSVSKPPGRKPAEKPLPKPRG 289

QY 104 NLPPRKR--AKTEDEKQRRVRLNRRAAQSSRRKRLEVEALEKKNKELETLLINQV 161
Db 290 RKPKPERPPSSSSSDSDSDVDRISEWKR-----RDEARRRLEARRRQEELRLREQ 345

QY 162 KTNLILVEELNRRSSSGVVTSS--SPLDSL-ODSITLSQQLFGSRDQ-----TMS 211
Db 346 EKE---EKERRRERADRGEAERGSGGSDLELREDDPEVKR---GRKGRGPPSSSDS 399

QY 212 NPEQSLMDQIMRSAANTVPNAPSLPPTSDKFEQTKEDDEQADDEMEQTHETK 271
Db 400 EPEALEEREAKKSARKP--QSSSTEPARKP-GQKEKRVPEKQQA-KPVKVERT-RKRS 454

QY 272 EAAAKEKNKSKQSRVSTD 289
Db 455 EGFSMDRKVEKKEPSVE 472
```

Job time : 46.1647 secs



Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension **.rup**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:05:10 ; Search time 193.805 Seconds
(without alignments)
1641.819 Million cell updates/sec

Title: US-10-663-450-2
Perfect score: 2294
Sequence: 1 MAFOQSSPLVKFEAPAESF.....QFQSGASHGCGDDGGIAGV 451

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2294	100.0	451	2	Q8TFP3_TRIRE	Q8tfp3 trichoderma
2	885	38.6	429	2	Q4HTT5_GIBZE	Q4htt5 gibberella
3	756	33.0	430	2	Q7SHF0_NEUCR	Q7shf0 neurospora
4	587	25.6	556	2	Q51KW8_MAGGR	Q51kw8 magnaporthe
5	437.5	19.1	350	2	Q8TFU8_EMENI	Q8tfu8 emericeella
6	426.5	18.6	342	2	Q6W8X2_ASFNG	Q6w8x2 aspergillus
7	424	18.5	347	2	Q5AQN3_EMENI	Q5aqn3 aspergillus
8	342	14.9	433	2	Q4WY88_ASFFU	Q4wey8 aspergillus
9	285	12.4	357	2	Q5AA52_CANAL	Q5aa52 candida alb
10	230.5	10.0	299	2	Q6CEV1_YARLI	Q6cev1 yarrowia li
11	211	9.2	230	1	HAC1_YEAST	P41546 saccharomyc
12	208.5	9.1	260	2	Q6B0C2_DEBHA	Q6bcq2 debaryomyc
13	206.5	9.0	273	2	Q6CKQ1_KLUUA	Q6ckq1 kluyveromyc
14	197	8.6	329	2	Q6FLY3_CANGA	Q6fly3 candida gla
15	190	8.3	228	2	Q75BQ5_ASHGO	Q75bq5 ashbya goss
16	172.5	7.5	1225	2	Q4P0Y9_USTWA	Q4p0y9 ustilago ma
17	172	7.5	595	2	Q4P8Q4_USTWA	Q4p8q4 ustilago ma
18	170.5	7.4	690	2	Q44743_CABEL	Q44743 caenorhabdi
19	169.5	7.4	588	2	Q411V8_GIBZE	Q411v8 gibberella
20	166	7.2	3257	2	Q9V736_DROME	Q9v736 drosophila
21	164.5	7.2	1444	2	Q9VTN2_DROME	Q9vtn2 drosophila
22	164.5	7.2	1514	2	Q8SY55_DROME	Q8sy55 drosophila
23	163.5	7.1	1196	2	Q4SUB7_TETNG	Q4sub7 tetraodon n
24	162	7.1	615	2	Q4WH07_ASFFU	Q4wh07 aspergillus
25	159.5	7.0	2938	2	Q4IR68_GIBZE	Q4ir68 gibberella
26	157.5	6.9	868	2	Q4P172_USTWA	Q4p172 ustilago ma
27	157	6.8	1214	2	Q4W3T0_ASFFU	Q4w3t0 aspergillus
28	156.5	6.8	782	2	Q870C5_9SACH	Q870c5 kluyveromyc
29	156.5	6.8	3340	2	Q4QB22_LEIMA	Q4qbb2 leishmania
30	155.5	6.8	2041	2	Q7S374_NEUCR	Q7s374 neurospora
31	155	6.8	583	1	API_KLUUA	P56095 kluyveromyc

RESULT 1									
Q8TFP3 TRIRE PRELIMINARY; PRT; 451 AA.									
ID	Q8TFP3	TRIRE							
AC	Q8TFP3								
DT	01-JUN-2002	(Tremblrel. 21, Created)							
DT	01-JUN-2002	(Tremblrel. 21, Last sequence update)							
DT	01-MAR-2004	(Tremblrel. 26, Last annotation update)							
DE	Transcription factor.								
GN	Name=hacl1								
OS	Trichoderma reesei (Hypocrea jecorina).								
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;								
OC	Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.								
OX	NCBI_Taxid=51453;								
RN	[1]								
RP	NUCLEOTIDE SEQUENCE.								
RA	Salohelmo M.L.A., Valkonen M., Penttilae M.E.;								
RL	Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.								
DR	EMBL;	A413272;	CAC88374.1;	-;	Genomic_DNA.				
DR	GO;	GO:0005634;	C:nucleus;	IEA.					
DR	GO;	GO:0003677;	F:DNA binding;	IEA.					
DR	GO;	GO:0006355;	P:regulation of transcription, DNA-dependent;	IEA.					
DR	InterPro;	IPR011700;	bZIP 2.						
DR	InterPro;	IPR004827;	TF_bZIP.						
DR	Pfam;	PF07716;	bZIP2; 1.						
DR	SMART;	SM00338;	BRIZ_1.						
DR	PROSITE;	PS50217;	BZIP; 1.						
KW	DNA-binding; Nuclear protein.								
SQ	SEQUENCE 451 AA; 49277 MW; 600F10B471EA3AD3 CRC64;								
Query Match 100.0%; Score 2294; DB 2; Length 451;									
Best Local Similarity 100.0%; Pred. No. 2.8e-104;									
Matches 451; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MAFOQSSPLVKFEAPAESF	LSAPGDNFTLSFADSTPSTLNPRDMTPSDVADISRLSV	60					
DB	1	MAFOQSSPLVKFEAPAESF	LSAPGDNFTLSFADSTPSTLNPRDMTPSDVADISRLSV	60					
QY	61	IPESQDAEDDESHSTSATAPSTSEKKPVKKRSGQVLPEPKTNLPKRKAKTDEKEQR	120						
DB	61	IPESQDAEDDESHSTSATAPSTSEKKPVKKRSGQVLPEPKTNLPKRKAKTDEKEQR	120						
QY	121	RVERVLNRAAQSSRRKRLVEALEKENKLETLLINVQKTNLLVVELNFRSSGV	180						
DB	121	RVERVLNRAAQSSRRKRLVEALEKENKLETLLINVQKTNLLVVELNFRSSGV	180						
QY	181	VTRSSPLDLSQDSTITLSQOLFSGRDGQTMSPQSLMDQIMRSAANPTVNPASLPSLP	240						
DB	181	VTRSSPLDLSQDSTITLSQOLFSGRDGQTMSPQSLMDQIMRSAANPTVNPASLPSLP	240						
QY	241	PISDKKEFTQKEDEEQADEDEMEQTHETKEAAAAAKENKQSRVSTSTQRPVSI	300						
DB	241	PISDKKEFTQKEDEEQADEDEMEQTHETKEAAAAAKENKQSRVSTSTQRPVSI	300						
QY	301	DAAPVFSDDAGANCLGLDPVHQDDGPFSGHISFGLSALDADRLYLLESQLLASPNASTV	360						

Q4IP09 gibberella
Q5KPM1 cryptococcu
Q20625 caenorhabdi
Q4P825 ustilago ma
Q9JJY5 mus musculu
Q569T3 xenopus lae
Q6C373 yarrowia li
Q9XEC3 bos taurus
Q54JY4 dictyosteli
Q61D33 caenorhabdi
Q80TR9 mus musculu
Q560M5 cryptococcu
Q6DE25 xenopus lae
Q6C621 yarrowia li

Db	301	DAAVPVFSDAGANCIGLDPVHQDGPFGSHGSLAALDADRYLLESQSLASPNASTV	360
Qy	361	DDLYLAGDSAACTNPLPDPDYDFDINDFTDDANHAAYDIIVAASNYAAADRELDLEIHD	420
Db	361	DDLYLAGDSNACFTNPLPDPDYDFDINDFTDDANHAAYDIIVAASNYAAADRELDLEIHD	420
Qy	421	ENQIPSRHSIIQQPQSGASSHGCDGGIAVGV	451
Db	421	ENQIPSRHSIIQQPQSGASSHGCDGGIAVGV	451
RESULT 2			
ID	Q4HTT5	GIBZE PRELIMINARY; PRT; 429 AA.	
AC	Q4HTT5;		
DT	13-SEP-2005	(T-EMBLrel. 31, Created)	
DT	13-SEP-2005	(T-EMBLrel. 31, Last sequence update)	
DE	13-SEP-2005	(T-EMBLrel. 31, Last annotation update)	
DE		Hypothetical protein.	
GN		ORFNames=FG11623.1;	
OS		Gibberella zeae PH-1.	
OC		Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	
OC		Hypocreales; Nectriaceae; Gibberella.	
OX		NCBI_TaxID=229533;	
RP		[1]	
RP		NUCLEOTIDE SEQUENCE.	
RC		STRATN=PH-1;	
RA		Birren B., Nussbaum C., Abouelleil A., Allen N., Anderson S.,	
RA		Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavskiy L.,	
RA		Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,	
RA		Choepeel Y., Collumore A., Cook A., Cooke P., Corum B., DeArellano K.,	
RA		Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,	
RA		Erickson J., Faro S., Ferreira P., FitzGerald M., Gage D., Galagan J.,	
RA		Gardyna S., Gnerre S., Graham L., Grand-Pierre N., Harez N.,	
RA		Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,	
RA		Jaffe D., Johnson R., Jones C., Kamal M., Kamat A., Karatas A.,	
RA		Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,	
RA		Ma L.-J., Mabbitt R., MacLean C., Macdonald P., Major J., Manning J.,	
RA		Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,	
RA		Mhova T., Mienga V., Murphy T., Naylor J., Nguyen C., Nicol R.,	
RA		Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neill D.,	
RA		Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,	
RA		Rachupka A., Ramaamy U., Raymond C., Retta R., Rise C., Rogov P.,	
RA		Roman J., Schauer S., Schupback R., Seaman S., Severy P., Smirnov S.,	
RA		Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,	
RA		Talamas J., Testaye S., Theodore J., Topham K., Travers M.,	
RA		Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,	
RA		Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,	
RA		Lander E.;	
RT		"Fusarium graminearum genome sequence."	
RL		Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.	
CC	-1-	CAUTION: The sequence shown here is derived from an	
CC		EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is	
CC		preliminary data.	
DR	EMBL;	AACM01000497; EAA78735.1; -; Genomic_DNA.	
KN		Hypothetical protein.	
SQ	SEQUENCE	429 AA; 46630 MW; 04CCDE40BD392901 CRC64;	
Query Match			
Best Local Similarity 38.6%; Score 885; DB 2; Length 429;			
Matches 215; Conservative 69; Mismatches 124; Indels 64; Gaps 12;			
Qy	1	MAPQSSPLVKFEAPSPAESFLSAPGDNFTSLPADSTP---STLPRDMPTPDSVADIDSR	57
Db	1	MEYQOPLDKFQNSPTESLMIPGDGFTSLPDVFTPSATSTMPFMENMTPKSYTD-DQI	59
Qy	58	LSVIPESQDAEDSHSTATAPSTSEKPKVKRSGQVLPEPKTNLP-----	107
Db	60	PSSLQPKKEED---MSTPSPGFA-PEKTKTKRSGQVLPEPKTNLPPTYGTLRHVV	115
Qy	108	-----RKRKTEDEKQRRVERVLNRRAAQSSRRERKLEVEALEKRNKELETLIN	159

```

Db 6 AHSPSPANKPE-NPASLLATPGEIPPEVPGSDRATSATPSLDVMPGVSLAD-DVDLTA 63
Qy 61 I-----PESQDAEDDESHSTSATAPSTSEKKPVKKRKGWQVLPEPKTNLPPPKRA 111
Db 64 LATLTPQIPPRSTPA-----STPAPET-EKKPVKKRKGWQVLPEPKTNLPPPKRA 114
Qy 112 KTEDEKQRRVRLNRRAAQSSRRKRLEVEALEKRNKELETLLINVKTNLILVEEL 171
Db 115 KTEDEKQRRVRLNRRAAQSSRRKRLEVEALEKRNKELETLLINVKTNLILVEEL 174
Qy 172 NRPRSSGVTRSSPLDLSQDS-ITLSQQLFGSRDGTMSNPSQSLMDQIMRSAANPTV 230
Db 175 RENGVAITATRPAS-FDGLNPTPTVTSQELFFSSQGHNSKHDSSLEQLFPYIKTETV 233
Qy 231 NPASLSPPLPIDSEKQTKKEEDEBEQADEMEQWTWHEKRAAAAKEKNSKQSRVSTDS 290
Db 234 NPASLSPVNLPLPEME-----EDGEKEQS---TAAQPVADATSTATVDTSPDA 278
Qy 291 TORPAVSIIGDAAVPVFSDDAGANCLGLDPVHQDGFSGHSGFLSAAALDADRYLLESQ 350
Db 279 TOHPAVVFPVAVSAPVAGSSLSAALGLGNFMDD-----LVLSKTFDADRYIYETE 329
Qy 351 LLASPNASTVDDVLAGDSAACTNPLPSVDYDFDINDFLTDDANHAAYDIVAASNYAAD 410
Db 330 FFSPPSPNSNDEVSMAAGDDETLHPANPENFPD--BFINPDVNSA-----ANSNEQOPQ 382
Qy 411 RE-----LDLEIHDPENQIPSRHSIQPOSGASSHCGDDGGIAGV 450
Db 383 QQSQFCATTNASVINLTPNASEDDPYQOPHTGASLNGCDGGIAGV 428

RESULT 4
Q51KW8 MAGGR
ID Q51KW8 MAGGR PRELIMINARY; PRT; 556 AA.
AC Q51KW8.
DT 13-SEP-2005 (TREMBLrel. 31, Created)
DT 13-SEP-2005 (TREMBLrel. 31, Last sequence update)
DE Hypothetical protein.
GN ORFNames=MG09010.4;
OS Magnaporthe grisea 70-15.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
OX NCBI_TaxID=242507;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Birren B., Nusbam C., Abebe A., Abouelleil A., Adekoya E.,
RA Alt-zahra M., Allen N., Allen T., An P., Anderson M., Anderson S.,
RA Arachi H., Armbruster J., Bachanteang P., Baldwin J., Barry A.,
RA Bayul T., Blitshsteyn B., Bloom T., Blye J., Boguslavskiy L.,
RA Boroweky M., Boukhgalter B., Brunache A., Butler J., Calixte N.,
RA Calvo S., Camarata J., Campo K., Chang J., Cheshatsang Y., Citroen M.,
RA Callimore A., Considine T., Cook A., Cooke P., Corum B., Cuomo C.,
RA David R., Dawoe T., Degray S., Dodge S., Dooley K., Dorje P.,
RA Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T., Engels R.,
RA Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,
RA Fitzgerald M., Foley K., Gage D., Galaqan J., Gearin G., Gnerre S.,
RA Girke A., Goyette A., Graham J., Grandbois E., Gyaltsen K., Hafez N.,
RA Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,
RA Jaffe D., Jones C., Kamal M., Kamat A., Kamvesselis M., Karlsson E.,
RA Kells C., Kieu A., Kiser P., Kodira C., Kulbokas E., Labutti K.,
RA Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,
RA Lindblad-toh K., Liu X., Lokytang T., Lokytang Y., Lucien O.,
RA Lui A., Ma L.J., Mabbitt R., Macdonald J., Maclean C., Major J.,
RA Manning J., Marabella R., Maru K., Matthews C., Maucell E.,
RA McCarthy M., McDonough S., McGhee T., Meldrum J., Menes L.,
RA Mesirov J., Mihalov A., Mihova T., Mikkelson T., Mlenga V., Moru K.,
RA Mozes J., Mulrain L., Munson G., Naylor J., Neues C., Nguyen C.,
RA Nguyen N., Nguyen T., Nicol R., Nielsen C., Nizzari M., Norbu C.,
RA Norbu N., O'donnell P., Okoawo O., O'leary S., Omotosho B.,

```

```

RA O'Neill K., Osman S., Parker S., Perrin D., Phunkhang P., Piquani B.,
RA Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond C.,
RA Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,
RA Rutman M., Schubach R., Seaman C., Settippalli S., Sharpe T.,
RA Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,
RA Spencer B., Stalker J., Stange-thomann N., Stavropoulos S.,
RA Stetson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,
RA Tenzing P., Tesfaye S., Theodore J., Thoulutseang Y., Topham K.,
RA Towey S., Tsamla T., Tsomo N., Vallie D., Vassiliev H.,
RA Venkataraman V., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,
RA Wangdi T., Whitaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,
RA Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembeck L.,
RA Zimmer A., Zody M., Zander E.;
RT "The genome sequence of Magnaporthe grisea.";
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Dean R., Mitchell T., Brown D., Pan H., Thon M.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Zhu H., Blackmon B.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
CC -I- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AACU01001606; BAA47880.1; -; Genomic_DNA.
DR InterPro; IPR011700; bZIP 2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07716; bZIP2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS0217; bZIP; 1.
DR DNA-binding; Hypothetical protein; Nuclear protein.
KW SEQUENCE 556 AA; 60798 MW; CFDC4C8D2DC4963C CRC64;
SQ
Query Match 25.6%; Score 587; DB 2; Length 556;
Best Local Similarity 45.0%; Pred. No. 6.6e-21;
Matches 148; Conservative 34; Mismatches 93; Indels 54; Gaps 9;
Qy 6 SSPLVKFEASPAESPLSAPGDNFTSLFADSTSTLNPRDMWTPDSVADIDSLSVI-PES 64
Db 6 STPELKMEASPAESPLSTFGDMYPSLFGDAKSFSLNPSDLASPGSQEDPVELDAKTES 65
Qy 65 QDAEDDESHST--SATAPSTSEKKPVKKRKGWQVLPEPKTNLPPPKRAKTEDEKQRRV 122
Db 66 AALRAVESTPTPEETSSPAAASEKTKKRSKGWQVLPEPKTNLPPPKRAKTEDEKQRRV 125
Qy 123 ERLVLRRAAQSSRRKRLEVEALEKRNKELETLLINVKTNLILVEELNRPFRSSGVVT 182
Db 126 ERLVLRRAAQSSRRKRQVEALEORNAALBQQLLHFQKLTQTMVQELQLRREAGVVT 185
Qy 183 RSSPLDLSQDSITLSQQLFGSRD-----GOTMSNPSQSLMDQIMRSAANP-TVNP 232
Db 186 SSEKP-----DGLTLPFLFRSDAIAASATSVSATLADPANSLED-IFRSTNVPTVNP 239
Qy 233 ASLSPLSPIDSKFOTKEEDEBEQADEMEQWTWHEKRAAAAKEKNSKQSRVSTDSQ 292
Db 240 ASIRSSPAP-----EQHESVASGEE-----AKTSADLTQ 269
Qy 293 RPAVSIIGD-----AAVPVFSDDAGANCL 316
Db 270 HPAEMLCITDLCQSAKVPQASSQMPASMM 298

RESULT 5
Q8TFUB EMENI
ID Q8TFUB EMENI PRELIMINARY; PRT; 350 AA.
AC Q8TFUB.
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)

```


RA	Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,	
RA	Choepl Y., Collymore A., Cook K., Cooke P., Corum B., DeArellano K.,	
RA	Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,	
RA	Ericksen J., Faro S., Ferreira P., FitzGerald M., Gage D., Galagan J.,	
RA	Gardyna S., Gnerre S., Graham L., Grand-Pierre N., Hafez N.,	
RA	Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,	
RA	Jaife D., Johnson R., Jones C., Kamal M., Kamat A., Karatas A.,	
RA	Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,	
RA	Ma L.-J., Mabbitt R., Maclean C., Macdonald P., Major J., Manning J.,	
RA	Matthews C., Mauceli E., McCarthy M., McDonald J., Meneus L.,	
RA	Mihova T., Miengva V., Murphy T., Naylor J., Nguyen C., Nicol R.,	
RA	Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neill D.,	
RA	Oliver J., Peterson K., Phunkhang P., Pierre N., Porcell S.,	
RA	Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,	
RA	Roman J., Schauer S., Schuback R., Seaman S., Severy P., Smirnov S.,	
RA	Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,	
RA	Talamas J., Tesfaye S., Theodore J., Topham K., Travers M.,	
RA	Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,	
RA	Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,	
RA	Zander E.;	
RT	"Genome Sequence of Aspergillus nidulans.";	
RL	Submitted (JAN-2004) to the EMBL/GenBank/DBJP databases.	
DR	EMBL; AACD01000172; EAA6464.1; -; Genomic_DNA.	
DR	GO; GO:0005634; C:nucleus; IEA.	
DR	GO; GO:0003677; F:DNA binding; IEA.	
DR	GO; GO:0006395; P:regulation of transcription, DNA-dependent; IEA.	
KW	DNA-binding; Hypothetical protein; Nuclear protein.	
SQ	SEQUENCE 347 AA; 37841 MW; A7CAAPEE761B3E1 CRC64;	
Query Match 18.5%; Score 424; DB 2; Length 347;		
Best Local Similarity 32.7%; Pred. No. 3.5e-13;		
Matches 147; Conservative 47; Mismatches 123; Indels 132; Gaps 20;		
Qy	26 DNFSTSL-----FADSTPTSLNPRDMWTPDSVADISRLSVIPESQDAEDDESHSTSATAP 80	
Db	5 DRFSFVKMEDAFANSLPT-----TPSLEVPV---LTVSPA-----DTSLOTKNVVA 47	
Qy	81 ST--SEKKPVKKRKSXGQVLPEPKTNLPKRKRAKTEDEKEQRRVVERVLNRRAAQSSRER 138	
Db	48 QTKPEKKPAKKRKSXGQELPVPKTNLPKRKRAKTEDEKEQRRIVRLNRFAAAQTSRER 107	
Qy	139 KRLEVEALEKRNKELETLLINVQTNLLVLELNFRFRSSGVVTRSSPLDSDLODSTLS 198	
Db	108 KRLEMKLESEK-----IDMEQNOQFLLRQAQMEANRLS-----QQVAQLS 151	
Qy	199 QQLFGSRDQGTMSNPEOSLMDQIMRSAANPTVNPASLSPSLPPISDKFQTKEDESQAD 258	
Db	152 AEVRGSRH-----STPTSS-----: : : : : -----EGDEVPLD 189	
Qy	259 EDEMEQTHETKEAAAAKERNKSKQSRVSTDSRTPAVSIGG---DAAVPVFSDDAGAN 314	
Db	190 RIP--FPTPSVTDVSYPTLKPS-----LADDVVRPAVSGVGLGEDSALTFL--DLGAS 239	
Qy	315 CLGLDPVHQDDGPPS-----IGHISFGLSALDADRYLLE-----SLLIASNASTVD 361	
Db	240 -IKHEPTHDLTAPISDDDFRLFNFGSDLSLEDGFAGFVLDSCDLSAFPFDMSVD 298	
Qy	362 DDYLAGDSAACTFNPPLPSDVFDFINDFTDDANHAADVAAASNYAAADRELDLDEIHDP 421	
Db	299 FD-----TBPVILE-DLEQTNGUSGASCKAASL-----: : : : : : ----- 326	
Qy	422 NQIPSRHS1QQPQSGASSHGCDGGIAGV 450	
Db	327 -----QPSHGASTSRCDGGIAG 345	
RESULT 8		
Q4WEY8 ASPFU		
ID	Q4WEY8 ASPFU PRELIMINARY; PRT; 433 AA.	
AC	Q4WEY8;	
DT	13-SEP-2005 (TrEMBLrel. 31, Created)	
DT	13-SEP-2005 (TrEMBLrel. 31, Last sequence update)	
DT	13-SEP-2005 (TrEMBLrel. 31, Last annotation update)	

DE	BZIP transcription factor (Haca), putative.
GN	ORFNames=Afu3g04070;
OS	Aspergillus fumigatus Af293.
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX	NCBI_TaxID=330879;
RN	[1]
RP	NUCLEOTIDE SEQUENCE.
RC	STRAIN=Af293;
RA	Nierman W., Pain A., Anderson M.J., Wortman J., Kim H., Stanley.,
RA	Arroya J., Brennan M., Abe K., Archer D.B., Bermejo C., Bennett J.,
RA	Bowyer P., Chen D., Collins M., Coulson R., Davies R., Dyer P.S.,
RA	Farman M., Fedorova N., Fedorova N., Feldblyum T.V., Fischer R.,
RA	Fosker N., Fraser A., Garcia J.L., Garcia M.J., Goble A.,
RA	Goldman G.H., Gomi K., Griffith-Jones S., Gwilliam R., Haas B.,
RA	Haas H., Harris D., Horiuchi H., Huang J., Humphrey S., Jimenez J.,
RA	Keller N., Khouri H., Kitamoto K., Kobayashi T., Kulkarni R.,
RA	Kumagai T., Lafton A., Lange J.-P., Li W., Lord A., Lu C.,
RA	Majoros W.H., May G.S., Miller B.L., Mohammad S., Molina M., Monod M.,
RA	Mounya I., Mulligan S., Murphy L., O'Neil S., Paulsen I.,
RA	Penalva M.A., Pettea M., Price C., Pritchard B.L., Quail M.A.,
RA	Rabinowitz E., Rawlins N., Rajandream M.-A., Reichard U.,
RA	Renauld H., Robson G.D., Rodriguez de Cordoba S., Rodriguez-Pena J.M.,
RA	Romling C.M., Rutter S., Salzberg S.L., Sanchez M.,
RA	Sanchez-Ferrero J.C., Saunders D., Seeger K., Squares R., Squares S.,
RA	Takeuchi M., Tekaja F., Turner G., Vazquez de Alandana C.R., Weidman J.,
RA	White O., Woodward J., Yu J.-H., Fraser C., Galagan J.E., Asai K.,
RA	Machida M., Hall N., Barrell B., Denning D.W.
RT	"Genomic sequence of the pathogenic and allergenic filamentous fungus
RT	Aspergillus fumigatus."
RL	Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
CC	!- CAUTION: The sequence shown here is derived from an
CC	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC	preliminary data.
DR	EMBL; AAHF01000010; EAL86689.1; -; Genomic_DNA.
DR	InterPro; IPR011700; bZIP_2.
DR	InterPro; IPR004827; TF_bZIP.
DR	Pfam; PF07116; bZIP_2; 1.
DR	SMART; SM00338; BRLZ; 1.
DR	PROSITE; PS50217; bZIP; 1.
DR	PROSITE; PS00036; bZIP_BASIC; UNKNOWN_1.
KW	DNA-binding; Nuclear protein.
SQ	SEQUENCE 433 AA; 48908 MW; 3742DD0B43E05C74 CRC64;
Query Match 14.9%; Score 342; DB 2; Length 433;	
Best Local Similarity 35.98; Pred No. 4.5e-09;	
Matches 99; Conservative 35; Mismatches 72; Indels 70; Gaps 10;	
Qy	26 DNFTSLFADSTPSTLNPRDMMTPDSVADIDSRLSVIPESQDAEDESHSTSATPSTSEK 85
Db	3 DNFASVVESLSGTSASALPLLT---VSPADTSLKA-PE-----TKVQETKTEEK 47
Qy	86 KPVKKRKSQGVLPPEKTNLPKRKATKTEDEKQRRVERVLNRNRAAQSRERKRLVEEA 145
Db	48 KPPKRRKSQGVQLPIPKTNLPKRKATKTEDEKQRRIVERLNRNRAAAQTSRERKRLMEK 107
Qy	146 LKKRKNKELTLLINVQKTNLILVELNRRRSRGVWTRSSSPDLSLQSDITLSQQLGCSR 205
Db	108 LENEK-----IQMEQNOQFLQRLSQMAENN-----RLSQQL----- 140
Qy	206 DGQTMNSPEQSLMDQIMRSAAFPV---NPASLSPSLPPIIDKFEQTKBEDE---EQADE 259
Db	141 -----AQLTAEVRRNSRNSTPKPGSPATASPTLTPTLFQK---EGDELPLERIPF 186
Qy	260 DEEMEQTHWTKEAAAKKKNQSKQSRVSTDSSTORPA 295
Db	187 PTPSITDYSPTLPKPSLLAE-----SSDVTQHPA 214
RESULT 9	
Q5AA52_CANAL	
ID	Q5AA52 CANAL PRELIMINARY; PRT; 357 AA.
AC	Q5AA52;

RESULT 8
Q4WEY8 ASPFU
ID Q4WEY8
AC Q4WEY8;
DT 13-SEP-
DT 13-SEP-
DT 13-SEP-

RESULT 10
Q6CEV1 YARLI


```

RESULT 11
HACL_YEAST
ID HACL_YEAST STANDARD; PRT; 230 AA.
AC P41546; P87040;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE HACL1 protein.
GN Name=HACL1; Synonyms=ERN4, IRE2; OrderedLocusNames=YPL031W;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=S288c;
RA MEDLINE=95116316; PubMed=7816617;
RA Nojima H., Leem S.-H., Araki H., Sakai A., Nakashima N., Kanaoka Y.,
RA Ono Y.;
RT "HACL1: a novel yeast bZIP protein binding to the CRE motif is a
RT multicopy suppressor for cdc10 mutant of Schizosaccharomyces pombe.";
RL Nucleic Acids Res. 22:5279-5288 (1994).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=S288c / AB972; PubMed=7670463;
RX MEDLINE=95400292;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Teuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M.-A., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae.";
RL Nat. Genet. 10:261-268 (1995).
RN [3]
RP SEQUENCE REVISION TO 183-230.
RA Murakami Y.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AB320;
RX MEDLINE=97222447; PubMed=9077435;
RA Mori K., Kawahara T., Yoshida H., Yanagi H., Yura T.;
RT "Signalling from endoplasmic reticulum to nucleus: transcription
RT factor with a basic-leucine zipper motif is required for the unfolded
RT protein-response pathway.";
RL Genes Cells 1:803-817 (1996).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=97086687; PubMed=8932376; DOI=10.1093/nar/24.21.4222;
RA Nikawa J.-I., Akiyoshi M., Hirata S., Fukuda T.;
RT "Saccharomyces cerevisiae IRE2/HACL1 is involved in IRE1-mediated KAR2
RT expression.";
RL Nucleic Acids Res. 24:4222-4226 (1996).
RN [6]
RP CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RX MEDLINE=97053779; PubMed=8998193; DOI=10.1016/S0092-8674(00)91360-4;
RA Cox J.S., Walter P.;
RT "A novel mechanism for regulating activity of a transcription factor
RT that controls the unfolded protein response.";
RL Cell 87:391-404 (1996).
CC -1- FUNCTION: Seems to be involved in the unfolded protein response
CC (UPR) pathway. Binds to the UPR element (UPRE) in the promoter of
CC UPR-regulated genes such as KAR2, PD11, EUG1 and PKB2; activates
CC the transcription of these genes.
CC -1- SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=the level of each isoform is regulated by a splicing
CC event that occurs when the UPR is induced by IRE1;
CC Name=U;
CC IsoId=P41546-1; Sequence=Displayed;
CC Name=I;

```

```

CC CC IsoId=P41546-2; Sequence=VSP_000586;
CC Note=Active and stable isoform which induces UPR;
CC -1- SIMILARITY: Belongs to the bZIP family.
CC -1- SIMILARITY: Contains 1 bZIP domain.
CC CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC CC
CC EMBL; D26506; BAA05513.1; -; Genomic DNA.
CC EMBL; D50617; BAA24425.1; -; Genomic DNA.
CC EMBL; D86413; BAA19565.1; -; Genomic DNA.
CC PIR; S78571; S78571.
CC GeneOnline; 140124; -.
CC TRANSFAC; T02039; -.
CC Ensembl; YPL031W; Saccharomyces cerevisiae.
CC SGD; S00001863; HAC1.
CC GO; GO:0005634; C:nucleus; TAS.
CC GO; GO:0003704; F:specific RNA polymerase II transcription fa...; TAS.
CC GO; GO:0003700; F:transcription factor activity; IDA.
CC GO; GO:0006644; P:phospholipid metabolism; TAS.
CC GO; GO:0045893; P:positive regulation of transcription, DNA-d...; IDA.
CC GO; GO:0006357; P:regulation of transcription from RNA polym...; TAS.
CC GO; GO:0006990; P:unfolded protein response, positive regulat...; TAS.
CC InterPro; IPR011700; bZIP 2.
CC Pfam; PF07716; bZIP 2; 1.
CC SMART; SM00338; BRLZ; 1.
CC PROSITE; PS00217; bZIP_BASIC; 1.
CC Activator; Alternative splicing; Complete proteome; DNA-binding;
KW Nuclear protein; Transcription; Transcription regulation;
KW Unfolded protein response.
FT DOMAIN 67 81 Leucine-zipper.
FT DNA_BIND 29 61 Basic motif.
FT VARSPLIC 221 230 AVITMTRKIQ -> EAQSLNSFELNDFPITS (in isoform I).
FT FTID=VSP_000586.
FT CONFLICT 143 230 RLCPRVCRFRVGRDFMGAAECLRRKMYQSRRLPYTIN
FT NLFDAVASPLADPLCDIDAGSLNLPDINSIDLNMWKNPAVIT
FT MTRKIQ -> ATLSPKSRDSDASDQETSWELQMFKNVPE
FT STTLPAVDNNLFDVASKRWQTHSATI (in Ref. 2).
FT RLCPRVCRFRVGRDFMGAAECLRRKMYQSRRLPYTIN
FT -> ATLSPKSRDSDASDQETSWELQMFKNVPESTTLPAV
FT DN (in Ref. 4).
FT SQ SEQUENCE 230 AA; 26582 MW; 43073BCCCC4709B CRC64;
Query Match 9.2%; Score 211; DB 1; Length 230;
Best Local Similarity 35.8%; Pred. No. 0.0055;
Matches 64; Conservative 26; Mismatches 51; Indels 38; Gaps 6;
QY 102 KTNLPKRKRAKTEDEKORRVERVLNRRAAQSRERKRLEVEALEKRNKELETLL--IN 159
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23 KSTLPPKRKRAKTEKORRIIRNRRAHQKREKRLHLQYLERKCSLLENLNSVN 82
QY 160 VQK-----TNLIIVEELNFR-----RSSGVVTRSSSPDLSQDSITLS----- 198
Db :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
83 LEKLADHEDALTCSDAFVASLDEYRDFQSTRGASLDTRASS--HSSSDTFTPSPLNCTM 140
QY 199 -----OQLFGSRDQGMSPESLMDQIMRSAANPTVN---PASLSPSLPDISD 244
Db :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
141 EPLRCRPRVCRFRVGRDFMGAAECLRRKMYQSRRLPYTINLFDVAVASPLADPLCD 199
RESULT 12
Q8BQC2.DEBHA
ID Q8BQC2.DEBHA PRELIMINARY; PRT; 260 AA.
AC Q8BQC2;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)

```

Similarities with ep[P41546 *Saccharomyces cerevisiae* HAC1 protein.
 OrderedLocusNames=DEHA0E07139g;
 Debaromyces hansenii (Yeast) (Torulasporea hansenii).
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Debaromyces.
 NCBI_TaxID=4959;
 [1]
 NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 STRAIN=ATCC 36239 / CBS 767;
 PubMed=1529992; DOI=10.1038/nature02579;
 Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
 Lafontaine I., de Montigny J., Marck C., Neugeglise C., Talla E.,
 Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
 Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykaesten C.,
 Boissane A., Boyer E., Cattolico L., Confanioleri F., de Daruvar A.,
 Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
 Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
 Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
 Nicoud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
 Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,
 Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
 Zenlou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
 Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,
 Wincker P., Souciet J.-L.;
 "Genome evolution in yeasts."
 Nature 430:35-44(2004).
 EMBL; CR382137; CAG87828.1; -; Genomic_DNA.
 GO; GO:0005634; C:nucleus; IEA.
 GO; GO:0003677; F:DNA binding; IEA.
 GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR DR
 GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR DR
 InterPro; IPR011700; bZIP 2.
 DR DR
 InterPro; IPR004827; TF bZIP.
 DR DR
 Pfam; PF07716; bZIP 2; 1.
 DR DR
 SMART; SM00338; BRLZ; 1.
 DR DR
 PROSITE; PS0217; bZIP; 1.
 DR DR
 PROSITE; PS00036; bZIP BASIC; UNKNOWN 1.
 DR DR
 Complete proteome; DNA-binding; Nuclear protein.
 KW
 SEQUENCE 260 AA; 29442 MW; 49B1819E3993BFF6 CRC64;
 SQ
 Query Match 9.1%; Score 208.5; DB 2; Length 260;
 Best Local Similarity 30.8%; Pred. No. 0.0083;
 Matches 65; Conservative 36; Mismatches 75; Indels 35; Gaps 6;
 QY 102 KTNLPKRAKTEDEKQRRVVERILNRRAAQSRRERKLEVEALEKRNKELETLINQV 161
 Db 14 KSSLPPKRAKTEKEQRRVERILNRRAAHAGREKRVHLYESVYLKEDNMRLO 73
 QY 162 KTNLILVEELNFRFRSSGVVTRSS-----PLDSLQDSITLSQQLFGSDGTMSN 212
 Db 74 S-----NFDGVCQMSAKQLAKIDMQEMDDLQD---LKDITNLNGSLAN 117
 QY 213 P--EGLMDQIMRSANPTVPASLPSLPISKEFQTKEEDEBEQADE-----EM 263
 Db 118 EGGDNDLDELTPPAEQDTRKRLSSANSKCSLTTLTDSVEQVSSDKVNQIPHIM 177
 QY 264 EQTWHTKEAAAKKSKQSRVSTDSTQRP 294
 Db 178 EPN-SDSNELLSITNTVNYLSPISINSP 207
 RESULT 13
 Q6CKQ1_KJUL4
 ID Q6CKQ1_KJUL4 PRELIMINARY; PRT; 273 AA.
 AC Q6CKQ1;
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Kluyveromyces lactis strain NRRL Y-1140 chromosome F of strain NRRL Y-
 DE 1140 of Kluyveromyces lactis.
 DE OrderedLocusNames=KJL40F08976g;
 GN Kluyveromyces lactis (Yeast).
 OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.

ON NCBI_TaxID=28985;

OX [1]

RY NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RZ STRAIN=CBS 2359 / IF0 1267 / NRRL Y-1140 / WM37;

RC PubMed=15229592; DOI=10.1038/nature02579;

RD Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,

RE Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,

RF Goffard N., Frangeul L., Aigle M., Anchaouard V., Babour A., Barbe V.,

RG Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,

RH Boisrame A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,

RI Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,

RJ Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,

RK Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller R.,

RL Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,

RM Pellenn S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,

RN Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,

RO Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,

RP Boucher C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,

RR Wunker P., Souciet J.-L.;

RS "Genome evolution in yeasts.";

RT Nature 430:35-44(2004).

RZ ENBL: CR382126; CAG98196.1; -: Genomic_DNA.

DR GO: GO:0005634; C:nucleus; IEA.

DR GO: GO:0003677; F:DNA binding; IEA.

DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR InterPro: IPR011616; bZIP 1.

DR InterPro: IPR004827; TF bZIP.

DR Pfam: PF001170; bZIP 1; 1.

DR SMART: SM00338; BRLZ; 1.

DR PROSITE: PS00036; bZIP_BASIC; 1.

DR Complete proteome; DNA-binding; Nuclear protein.

KW SEQUENCE 273 AA; 30429 MW; 97DFDFB1F8ACF5 CRC64;

QY Query Match 9.0%; Score 206.5; DB 2; Length 273;

Best Local Similarity 32.4%; Pred. No. 0.011;

Matches 77; Conservative

Qy 102 KTNLPPRKRAKTEDEKEQRVERVLNRRAAQSRERKRLVBALEKRNKELETLINQV 161

Db 15 KPTLPKRAKTEDEKEQRRIILNRRAHQSRERKRLVORLEEKCHLLEGIL---K 71

Qy 162 KTNLILVEILNRRFRSSGVTR-----SSPLDSLSDSLTSLQQLFGSRDQGTMSNPEQ 215

Db 72 MVDLIDILSENN--AKLUGMYEQWREMQVDSGSISSHSNT--GMLDSPESLT-SSPDK 125

Qy 216 SLMDQIMRSANPTVNPASLSPS---LPPISDKEFQTKEDDEEQADEEMEQTWHETKE 272

Db 126 K--DHYSHSHSRISISSSSSSPSNLP-----HGMVTDNGMLDED----- 164

Qy 273 AAAAKEKNS-----KQSRVSTDTQRPAVSGIDAAVPVPSD-DAGANCLGLDPHQD 324

Db 165 -----NNSLNYILQQNYQLSST--PVMKLEEDSHMLLENNGDADLNDVGISFAED 214

RESULT 14

Q6FLY3 CANGA

ID Q6FLY3 CANGA PRELIMINARY; PRT; 329 AA.

AC Q6FLY3;

DT 05-JUL-2004 (TrEMBLrel. 27, Created)

DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)

DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)

DE Candida glabrata strain CBS138 chromosome K complete sequence.

GN OrderedLocusNames=CAGL0K12540g;

OS Candida glabrata (Yeast) (Totulopsis glabrata).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; mitosporic Saccharomycetales; Candida.

ON NCBI_TaxID=5478;

OX [1]

RY NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RZ STRAIN=ATCC 2001 / CBS 138;

RC PubMed=15229592; DOI=10.1038/nature02579;

RD Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.

RE

RA	Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,	RT	"The Ashbya gossypii genome as a tool for mapping the ancient
RA	Goffard N., Frangeul L., Aigle M., Anthoud V., Babour A., Barbe V.,	RT	Saccharomyces cerevisiae genome.";
RA	Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykaaten C.,	RL	Science 304:304-307(2004).
RA	Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,	DR	EMBL; AB016816; AAS51442.1; -; Genomic_DNA.
RA	Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,	DR	AGD; ACR216C; -.
RA	Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,	DR	GO; GO:0005634; C:nucleus; IEA.
RA	Kerrest A., Kozul R., Lemaire M., Lesur I., Ma L., Muller H.,	DR	GO; GO:0003677; F:DNA binding; IEA.
RA	Nicaud J.-M., Nikolski M., Ottas S., Ozier-Kalogeropoulos O.,	DR	GO; GO:0003655; P:regulation of transcription, DNA-dependent; IEA.
RA	Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,	DR	InterPro; IPR011700; bZIP 2.
RA	Swennen D., Tekia F., Wesolowski-Louvel M., Westhof E., Wirth B.,	DR	InterPro; IPR004827; TF bZIP.
RA	Zenou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,	DR	Pfam; PF07716; bZIP 2; 1.
RA	Bouchier C., Caudon B., Scarpelli C., Gaillardin C., Weissenbach J.,	DR	PROSITE; PS00036; BZIP_BASIC; 1.
RA	Wincker P., Souciet J.-L.,	KW	Complete proteome.
RT	"Genome evolution in yeasts.";	SQ	SEQUENCE 228 AA; 24914 MW; 605A60D78633F038 CRC64;
RL	Nature 430:35-44(2004).		
DR	EMBL; CR30957; CAG61724.1; -; Genomic_DNA.		
DR	GO; GO:0005634; C:nucleus; IEA.		
DR	GO; GO:0003677; F:DNA binding; IEA.		
DR	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.		
DR	InterPro; IPR011616; bZIP 1.		
DR	InterPro; IPR004827; TF bZIP.		
DR	Pfam; PF00170; bZIP 1; 1.		
DR	SMART; SM00338; BRLZ; 1.		
DR	PROSITE; PS0217; bZIP; 1.		
DR	PROSITE; PS00036; BZIP_BASIC; 1.		
KW	Complete proteome; DNA-binding; Nuclear protein.		
SQ	SEQUENCE 329 AA; 37162 MW; 469614BF3AF2282 CRC64;		
		Query Match	8.6%; Score 197; DB 2; Length 329;
		Best Local Similarity	23.2%; Pred. No. 0.039;
		Matches	66; Conservative 46; Mismatches 95; Indels 78; Gaps 6;
QY	102 KTNLPKRAKTEDEKEQRRVRLNRRAAASSRERKLEVEALEKRNKELETLLINQV 161		
Db	26 ETWMPKRAKTKKEEIKRIQLNRKAAQKSRDRKRYVANKKTKTKV----- 80		
QY	162 KTNLLIVELNFRSSGVVTRSSSLDLSITLSQQLFGSRDGTMSNPESQSLMDQI 221		
Db	81 -----LDQLQSKIDIKSMILDPFVVDVTYINMEQD--DEM 112		
QY	222 MRSAANTVNPASLSPSPSPKEQTEKEDEEQADEEEMEQTHETKEAAAKERKS 281		
Db	113 SFSSDTPSPSTTEISTLAENAVDVHAAEDVDNREYISSMIQ-----KKST 160		
QY	282 KOSRVSTDSQTPAVSIGGDAAPVPSDDAGANCIGLDPVHODGDFSGHSGFLSAALD 341		
Db	161 KFGSIGSDVTSTPVRPRSTEQMTPLTASTSSSTC-----NSAVIS 200		
QY	342 ADRYLLESQLLASPNASTVDDYLAGDSAACTNPLPSDYDFDIN 386		
Db	201 ASDNEVDKR-----QTITNPLSSAST-----PNKYDLRIN 232		
RESULT 15			
Q75B05	ASHGO		
ID	Q75EQ5_ASHCO PRELIMINARY; PRT; 228 AA.		
AC	Q75EQ5;		
DT	05-JUL-2004 (TremBLrel. 27, Created)		
DT	05-JUL-2004 (TremBLrel. 27, Last sequence update)		
DT	05-JUL-2004 (TremBLrel. 27, Last annotation update)		
DE	ACR216Cp.		
GN	Name=ACR216C;		
OS	Ashbya gossypii (Yeast) (Eremothecium gossypii).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Eremothecium.		
OX	NCBI_TaxID=33169;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].		
RC	STRAIN=ATCC 10895;		
RX	PubMed=15001715; DOI=10.1126/science.1095781;		
RA	Dietrich F.S., Voegeli S., Brachat S., Lerch A., Gates K., Steiner S.,		
RA	Mohr C., Poehmann R., Luedi P., Choi S., Wing R.A., Flavler A.,		
RA	Gaffney T.D., Philippsen P.;		

Search completed: November 23, 2005, 03:24:56
Job time : 197.805 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:07:01 ; Search time 31.6947 Seconds
(without alignments)
1369.117 Million cell updates/sec

Title: US-10-663-450-2
Perfect score: 2294
Sequence: 1 MAPQSSPLVKFEAPAESF.....QPQSGASSHGCDGGIAGV 451
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	242.5	10.6	238	2 S78571	transcription fact
2	170.5	7.4	688	2 T2750	hypothetical prote
3	154	6.7	962	2 T2459	hypothetical prote
4	153.5	6.7	2062	2 A59297	myosin X - mouse
5	153.5	6.7	5327	2 T13564	microtubule-associ
6	151	6.6	2484	2 T26216	hypothetical prote
7	151	6.6	2607	2 T26215	hypothetical prote
8	148.5	6.5	650	2 S16706	transcription acti
9	148.5	6.5	5170	2 T15348	hypothetical prote
10	147.5	6.4	2058	2 A59267	myosin X - human
11	146	6.4	1302	1 JC6009	surface-located me
12	145	6.3	2052	2 T18519	myosin X - bovine
13	144	6.3	1002	2 S70292	FUN12 protein - ye
14	143.5	6.3	168	2 T50922	bZIP protein HY5 l
15	143.5	6.3	326	2 T08591	TGACG-motif bindin
16	143	6.2	406	2 S38170	SRP40 protein - ye
17	142.5	6.2	322	2 T08592	TGACG-motif-bindin
18	141.5	6.2	1365	2 T30822	lmp1 protein - Myc
19	140.5	6.1	381	2 S26812	transcription fact
20	140.5	6.1	753	2 F96781	hypothetical prote
21	140	6.1	544	2 T40058	probable chromatin
22	140	6.1	1621	2 A82255	hypothetical prote
23	139	6.1	1132	2 T43483	translation initia
24	139	6.1	2722	2 T20532	hypothetical prote
25	138.5	6.0	1142	2 S59359	GIN4 protein - yea
26	138	6.0	619	2 T08613	hypothetical prote
27	138	6.0	624	2 PC6003	surface membrane p
28	138	6.0	768	2 H54024	protein kinase (EC
29	138	6.0	2526	2 T20531	hypothetical prote

30	138	6.0	2738	2 E88320	protein F07A11.6 l
31	137.5	6.0	384	2 G86287	hypothetical prote
32	137.5	6.0	582	2 I50224	erythroid cell tra
33	137.5	6.0	669	2 JC5662	hepatoma-derived g
34	137.5	6.0	777	2 F54024	protein kinase (EC
35	137	6.0	532	2 T06029	hypothetical prote
36	136.5	6.0	706	2 T49700	related to Ap-1-li
37	136.5	6.0	783	2 A58817	cyclin-dependent k
38	136	5.9	1337	2 T30291	dextranase - Strept
39	135.5	5.9	1824	1 QRHUMT	microtubule-associ
40	135	5.9	412	2 S86203	hypothetical prote
41	135	5.9	3924	2 S37431	ankyrin 2, neuropa
42	134.5	5.9	720	2 T26819	hypothetical prote
43	134	5.8	581	2 T24455	hypothetical prote
44	134	5.8	611	2 T22456	hypothetical prote
45	134	5.8	714	2 T22454	hypothetical prote

ALIGNMENTS

RESULT 1

S78571
transcription factor HAC1 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YFL031w
C;Species: Saccharomyces cerevisiae
C;Date: 13-Sep-1998 #sequence revision 13-Sep-1998 #text_change 09-Jul-2004
C;Accession: S78571; S56223; S53578
R;Murakami, Y.
submitted to the Protein Sequence Database, January 1998
A;Reference number: S78570
A;Accession: S78571
A;Molecule type: DNA
A;Residues: 1-238 <MUR>
A;Cross-references: UNIPROT:P41546; UNIPARC:UPI0000157F61; EMBL:D50617; MIPS:YFL031w
A;Note: this is a revision to the sequence from reference S56186
R;Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sasanuma
submitted to the EMBL Data Library, May 1995
A;Description: Analysis of the nucleotide sequence of chromosome VI from Saccharomyces cerevisiae
A;Reference number: S56186
A;Accession: S56223
A;Molecule type: DNA
A;Residues: 1-191, 'RWQTHSATI', <MUM>
A;Cross-references: UNIPARC:UPI0000179B60; EMBL:D50617; NID:G836685; PID:G836723; MIPS:YFL031w
A;Note: this sequence has been revised in reference S78570
A;Note: this was believed to be the complete sequence of YFL031w
R;Nojima, H.; Leem, S.H.; Araki, H.; Sakai, A.; Nakashima, N.; Kanaoka, Y.; Ono, Y.
Nucleic Acids Res. 22, 5279-5288, 1994
A;Title: Hac1: a novel yeast bZIP protein binding to the CRE motif is a multicopy suppressor of
A;Reference number: S53578; MUID:95116316; PMID:7816617
A;Accession: S53578
A;Molecule type: DNA
A;Residues: 1-142, 'RLCRPRVCRFRVGRDFMGAAECRLRRKMYQSRRLVPTYI', 183-220, 'AVITWTRKLQ' <NOJ>
A;Cross-references: UNIPARC:UPI000012C042; GB:D26506; NID:G633122; PIDN:BAA05513.1; PID:G633122
C;Genetics:
A;Gene: SGD:HAC1; IRE2; ERN4
A;Cross-references: SGD:S0001863; MIPS:YFL031w
A;Map position: 6L
A;Introns: 221/1
C;Keywords: DNA binding; nucleus; transcription factor

Query Match	10.6%	Score	242.5	DB 2	Length	238			
Best Local Similarity	31.4%	Pred. No.	8.7e-07						
Matches	76	Conservative	30	Mismatches	57	Indels	79	Gaps	9
Qy	102	KTNLPPRKAETDEKEQRRVERVLNRRRAAQSRRERKLEVEALEKRKNKELETL--IN	159						
		: :							

```
QY 205 RDQOTMSNEQSLMDQIMRASAANTVNPASLSPSPISDKFQTKEDDEQADEEME 264
Db 125 -----HSSSDTFTSPINCIMEPATLSP-----KSMRDSASDQE 158
QY 265 QTWHTKEAANAARAKKSKRSRSTSTORPAY-----SIGDDAAVPVFSDDAGANCLG 317
Db 159 TSW-----ELQMPKTEN-----VPSESTTLPAVDNNNLFDAVASPLADPLCDDIAG-NSLP 207
QY 318 LD 319
Db 208 FD 209
RESULT 2
T32750
hypothetical protein F57B10.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession: T32750
R:Greco, T.; Elliott, G.; Keppeler, D.
submitted to the EMBL Data Library, December 1997
A:Description: The sequence of C. elegans cosmid F57B10.
A:Reference number: Z21219
A:Accession: T32750
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-688 <GRE>
A:CROSS-references: UNIPROT:O44743; UNIPARC:UPI0000079D06; EMBL:AF039713; PIDN:AAB96719.
A:Experimental source: strain Bristol N2; clone F57B10
C:Genetics:
A:Gene: CESP:F57B10.1
A:Map position: 1
A:Introns: 54/1; 150/3; 196/1; 306/3; 383/3; 627/3
Query Match 7.4%; Score 170.5; DB 2; Length 688;
Best Local Similarity 21.8%; Pred. No. 0.031;
Matches 94; Conservative 63; Mismatches 125; Indels 149; Gaps 17;
QY 4 QOSSPLVAFEAS-----PAESFLSAGDNFTSLFADSTPFLNPRDMWTPDSVADIDS 57
Db 187 QNSGLVRFKSKQPRVLPNPASISLNAPSSFNQSTSTPAT----- 228
QY 58 LSVIPESQDAEDDESHSTSAT-----APSTSEKK-----PVKKRKSQGVLPPE-- 100
Db 229 -----SSSSSSTNGGFKSGSTGERRKVPPLRLDEESIKLKCKEGICLPDPFF 275
QY 101 PXTNLPPRKRAKTEDEKEQRRVERVLNRRRAQSSRRKRLEVEAL-----KQNKEL 153
Db 276 PLTK-----AERDLKRIRKIRNKRSQAOTSRKRKQDYIEQLERVSSESTKENQAL 326
QY 154 ETLLINVKTNLILVEELNRRSSGVVTRSS-----PLDSLQDSITLSQOL--FGSR 205
Db 327 KQOIERLSENOQSVISQLKQALQGNAKRTTQAGRCGLAVFMLSACLVLVSPQLSPLGNQ 386
QY 206 DQQTW-----SNPEQSLMD-----QIMRSA 225
Db 387 DNQKVLCEIEACQFSATSMNSANSQAIAIGVTAAPSVVIPSGGPMVSTNANRQWNRNA 446
QY 226 A-----NPTVNPASLSPSPISDKB-----FQTKEDDEQADEEMEQTWHTKEAA 274
Db 447 VLNHHNNSKYPASGNQNHPIALEDLNHPPTLPQKSYQQQ--HOPSMYRSDTIAMA 504
QY 275 AAKKXKSKRSRSTSTORPAYSIGDDAAV-PVF-----SDAG-ANCL 316
Db 505 MAKIGARKGSSTSSASSVASSTSTSSATSTPIYRTLTGLAFEDQCDASSDDSCANNP 564
QY 317 GLDPVHQDDGP 327
Db 565 SLVPKMSAQP 575
```

RESULT 3
T22459

```
hypothetical protein F49E2.5b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T22459
R:Sturston, J.
submitted to the EMBL Data Library, October 1994
A:Reference number: Z19566
A:Accession: T22459
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-962 <WIL>
A:CROSS-references: UNIPROT:Q20625; UNIPARC:UPI000017B9BF; EMBL:Z46267; PIDN:CAA86429.1;
A:Experimental source: clone F49E2
C:Genetics:
A:Gene: CESP:F49E2.5b
A:Map position: X
A:Introns: 37/2; 66/3; 210/3; 284/3; 469/3; 499/3; 522/3; 585/3; 637/3; 779/1; 825/3; 866/3
Query Match 6.7%; Score 154; DB 2; Length 962;
Best Local Similarity 20.5%; Pred. No. 0.38;
Matches 88; Conservative 77; Mismatches 164; Indels 100; Gaps 18;
QY 59 SVIPESQDAEDDESHST--SATAPSTSEK-KPVKKRKSQGVLP-----PEPKTNLPKRKA 111
Db 41 AVLPNISINDQRKEAFASLELTEPQQPOQVEKVKSEKKAQKQIAKHDEARQKVN-----A 94
QY 112 KTEDEKQRRVERVLNRRRAQSSRRKRLEVEAL----EKRNKELETLLINVKTNLIL 167
Db 95 KKAARKEARRAEAEAKKRAQEEHKKQWKAQERIKQEQERKEADLKQLQAEKKKEKAVK 154
QY 168 VEELNRRFRSSGVVTRSSPLDSLQDSITLSQQLFGSRDGGQTMSPNEQSLMDQIMRSAAN 227
Db 155 AEKAEKAEK-----TKKASTPAPVEEIVVKV-----ANDRSAAPAEPEK-----T 196
QY 228 PTVNPASLSPSPISDKFQTKEDDEQ-----ADEDEMEQ---TWHTKEAANAARKE 278
Db 197 PTNTPAEPAEQVQIEITGKKNKKNKSESEATAAPASVEQVQKPVVTEEPHQQAAPQE 256
QY 279 KNSKQSRVSTSTQRPAYSIGDDAAVPVFSDDAGANCLGLDPVHQDDGPFSS----- 329
Db 257 KKNKKRKSSENVPAAS-----ETPV-----EPVVENESPESEPVAKLITV 299
QY 330 -----TGHSGFLSALD-----ADRYLLESQSLASPNASTVDDDDYLAGDSAACTFN 375
Db 300 SNTEASAVNVMGFSDIVTPKADDEVITQDPVSAKQVLPVHPVSEIPEEPA-----VS 352
QY 376 PLPSDYDFDINDFLT-----DDANHAAYDIVAASNYAAADRELDLEIHDPE---NQI-PS 426
Db 353 KKPTADSMDFLDFVTPKTEAESTSEAPAPVWVSKPTESIEDLEIVTYEHVADVGTGNTLSPS 412
QY 427 RHSIQQPOS 435
Db 413 QHSTPSPNS 421
RESULT 4
A59297
myosin X - mouse
C:Species: Mus musculus (house mouse)
C>Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jul-2004
C:Accession: A59297
R:Yonezawa, S.; Kimura, A.; Koshiba, S.; Masaki, S.; Ono, T.; Hanai, A.; Sonta, S.; Kagei, Y.
Biochem. Biophys. Res. Commun. 271, 526-533, 2000
A>Title: Mouse myosin X: Molecular architecture and tissue expression as revealed by north-south hybridization
A:Reference number: A59297
A:Accession: A59297
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-2062 <YON>
A:CROSS-references: UNIPROT:Q9JJY5; UNIPARC:UPI00000289A0; GB:AJ249706; PIDN:CAB56466.2
A:Experimental source: strain BALB/c; tissue lib testis
C:Genetics:
A:Gene: myo10
```

F;66-727/Domain:: myosin motor domain homology <MMO>

```
Query Match          6.7%; Score 153.5; DB 2; Length 2062;
Best Local Similarity 22.8%; Pred. No. 1;
```

Qy		91	RKSWGQVLPBPKTNLPPKR-----AKTEDEKSORRVERVLNRRAAQSRRERKLEVEA	145
Dd		805	RVYROLLAB-KRELEBKREBEKEEERERERAQREADLLRAHQEAFTRRQOELE	863
Qy		146	LEKNKELETLLINVOKTNLILVELANFRSSGV--VTRSSSPLDLSIQDSITLSQQQLFG	203
Dd		864	LQKSQAAD-----LTRELKQRNKQVEIILRLKEXIEDLQ-----R	901
Qy		204	SRDQTMSNPESQLMDQIMRSAANTVPNASLPSLPISDKFQTKEEDEEQADEEM	263
Dd		902	MKERQELSTEASLQ-----KLQLRDEULRLLEDEACRAAQSFLE	942
Qy		264	EOTWHTYEBAANAKEKN-SKQSRVSTSTORPAVSIGDDAAVPVF-----	307
Dd		943	SLNFDEIDECVRNIERSLVSGSEIBELGELAESAAGEK--PSFNFSQPYPPEEEVDEG	1000
Qy		308	--SDDACANCLGLDPVHODDGPFISGHISFGLSAAL-----DADRVLLESOLLASPNA-	357
Dd		1001	FEADD-----DAFKDSPNPSBHSHSDORTSGIRTSDESSESSEDPMNDVTVPSTPSAD	1052
Qy		358	STV-----DDOYL-----AGDSAACF-TNP--LPS---DYDFINDFLTDANHAAYDI	400
Dd		1053	STVLLAASMQDSASLNHSSGESYICPNQPGDLPSPDGYDYDQDDY-EDGAITSGSVV	1111
Qy		401	VAAASNYPAA	408
Dd		1112	TFFNSNYGS	1119

RESULT 5
Tl3564
microtubule-associated protein homolog - fruit fly (*Drosophila melanogaster*)
N/Alternate names: hypothetical protein EG:49E4.1
C/Species: *Drosophila melanogaster*
C/Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 31-Dec-2004
C/Accession: Tl3564
R/Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
A/Submitted to the EMBL Data Library, April 1999
A/Description: Sequencing the distal X chromosome of *Drosophila melanogaster*.
A/Reference number: 217689
A/Accession: Tl3564
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-5327 <SPA>
A/Cross-references: UNIPROT:O76891; UNIPARC:UPI0000110149; EMBL:AL031128; PIDN:CAA20006
C/Genetics:
A/Cross-references: FlyBase:FBgn0025392
A/Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A/Note: EG:49E4.1

Query Match	6.7%	Score 153.5	DB 2	Length 5327
Best Local Similarity	20.4%	Pred. No. 3.2		
Matches	98	Conservative 84	Mismatches 163	Indels 135
			Gaps 21	
Qy	68	EDDESHSTATAPSTSEKPKVKRKSNGQVLPEPKTNLPKR	-----AKT-EDEK	117
Db	4140	EEDEHRFS--PPSDVDKAAI	-----TPQMPRLSPREEVAKIVADVAKVLSKD	4189
Qy	118	-----EQRRVERVLRN-----RRAAOSSRRERKEVEA-----LEKRNKELETLI	158	
Db	4190	DITDIIPDFBERQLEELKLTADTEESSDKSTRDEKSLIETSVKVEISEKSSPD-----	4243	
Qy	159	NVQKTNLILVELNRRFRSSGVWTR-----SSSP--LDSIQDSITLSQLFGSRDGTMS	211	
Db	4244	--QKSGPISIEEKDKIKQSEKAQURQIGILASSRPESVASQSPSPQSAASHEHKEVE	4301	
Qy	212	NPQSQLMDQIMR--SAANPTVNPASLSPSLAPIISDKFEQTKEEDBEQAD-----	258	

```

Db      4302  LSSHKAEKSSRPESVASQSEKDKMTKSPASSTQSFSTKEGDEETESLLHSLTTTETV 4361
Qy      259  EDMEQO-----TWHTKEAAAAKKNKSKQSRVST-----DSTQRPAVSI-- 298
Db      4362  ETQMEKKSFSFVSSTVTKTVLSQSQTQVLRREBTSSESLSLKLKVEDSSRRESLSLL 4421
Qy      299  ---GGDANVPVSDDAGANCLGLDPV---HDDGPFSGHSGFLSAA-----LDADRY 345
Db      4422  AERGGIATNTSLKEDTSSASQLEELLVQSECSSESIVSEIQTSIAQKSNKEIKDARET 4481
Qy      346  LLESQLLASPNASTVDDDDYLAGDSAACTNPLPSDYDFDINDPLTDANHAAYDIV--- 401
Db      4482  KVTSQFTTTSSAYKQDS--LKETVAEP---LATEKIVSAKEAFSTEAATKSADDCLKKT 4536
Qy      402  --AASNYAAADREL-----DLETHDPENQIPSRHSIIQOPOS 435
Db      4537  ASAVSSTSASQALFVCTGDSRRRESLLSQASESRLTHSDPEDEPADDVDRSSVKESRS 4596

RESULT 6
T26216
hypothetical protein W06A7.3c - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T26216
R:Ainscough, R.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z20173
A:Accession: T26216
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2484 <WIL>
A:Cross-references: UNIPROT:Q9U347; UNIPARC:UPI00000827B8; EMBL:Z78066; PIDN:CAL
A:Experimental source: clone W06A7
C:Genetics:
A:Gene: CESP:W06A7.3c
A:Map position: 5
A:Introns: 1827/1, 1866/3; 1963/3; 1990/3; 2262/1; 2289/1; 2339/2; 2463/2

Query Match          6.6%; Score 151; DB 2; Length 2484;
Best Local Similarity 21.3%; Pred. No. 1.8;
Matches 111; Conservative 66; Mismatches 189; Indels 156; Gaps 23;

Qy      37  PSTLNPRDMTPDSVADIDSRLSVIP-----ESQDAEDDESHSTSATAPSTSEKK 86
Db      1644  PKYAKPLD-----DSKSRVFAPLNIKLGRTYSEEQOKELVESLERPLTI-ITQKP 1694

Qy      87  PVKRRKSWGVLPEPKTNLPKRKATED-----EKEQRRVERVURNRAAQ 133
Db      1695  PERPTEDIGALSP-----LSPNTLAERYEVPVMDMQSVPHSPQEKQEEIALSEIIEEPQ 1749

Qy      134  SSERKRLEVEALEKENKLETLINVKQTNILVE-----ELNRRF----- 175
Db      1750  AMKEVPKPVESAPEKONESLEAPEINPIRVLVETKIMGPKSLNEDNDDDDGSECL 1809

Qy      176  -----RSSGVVTRSSPLD--SLQDSITLSQQLFGSRD--GQTMN-----PEQSLMDQ 220
Db      1810  DSTGDLSSERTIQRPNTSIDDPRIRRDSFSSISFGDRQKFRATENIRODLLPFQSSVSG 1869

Qy      221  IMRSAANPTV-----NPASLSPLSPLPI---SDKEFQTK--BEDEQAD---EDBE 262
Db      1870  YLRSSPNPSQQLLVNTLSMDSPSDLSPNAPPVGFNTAQFLKQLQEDRPSAEGSIDSSG 1929

Qy      263  MEQTWHTKEAAAAKKNKQSRVSTDSQRPASVITGGDANVPVSDDAGANCLGLDPVH 322
Db      1930  FEKVDHEGLDEFAAPVH-----DPMQK---SVFGSLG---SDD-----MKPGS 1967

Qy      323  QDDGPFSGHSGFLSALDADRYLLESQLLASPNASTVDDDDYLAGD--SAACFTNPLPS- 379
Db      1968  QDDGPFVIERNEANEATLK-----KNQKMSHHNDVIEKNVFNQNPATAALLSEPIASE 2021

Qy      380  -----DYDFDINDPLTD-----DANHAAYDIVAASN 405

```

Db	2022	ARKLVQDAVESAYKKQAVDSGEIGRELLDNVEQKIBQVKPEIVDSLHKAYDVGVDV	2081
Qy	406	YAAADRELDLEIHPENQIPSR---HSIQOQSGASSHGCD	444
Db	2082	HETVPNAVDDFVREAEKQLPESVPPEKIETPPLVDIHTVD	2123
RESULT 7			
T26215			
hypothetical protein W06A7.3a - Caenorhabditis elegans			
C:Species: Caenorhabditis elegans			
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004			
C:Accession: T26215			
R:Ainscough, R.			
submitted to the EMBL Data Library, August 1996			
A:Reference number: Z20173			
A:Accession: T26215			
A>Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-2607 <WIL>			
A:Cross-references: UNIPROT:Q23187; UNIPARC:UPI000007623E; EMBL:Z78066; PIDN:CAB01522.2;			
A:Experimental source: clone W06A7			
C:Genetics:			
A:Gene: CESP:W06A7.3a			
A:Map position: 5			
A:Introns: 1827/1; 1866/3; 1963/3; 1990/3; 2262/1; 2289/1; 2412/1; 2462/2; 2586/2			
Query Match			
Best Local Similarity 6.6%; Score 151; DB 2; Length 2607;			
Matches 111; Conservative 66; Mismatches 189; Indels 156; Gaps 23;			
Qy	37	PSTLPRDMWTDPDVADIDSLSVIP-----ESODADESHSATSATPSTSEKK	86
Db	1644	PKVAPLD-----DSKSRVFAPLNTKLGRTYSEEQKELVLSLERPLTI-ITQXP	1694
Qy	87	PVKGRKSGWQVLPEPKTNLPFRKRAKTED-----EKEQRRVERVLRNRAAQ	133
Db	1695	PEKPTEDIGALSP----LSPNTLAEYEVPMWQSVPHSQEKQEEIALSEIIEEPQ	1749
Qy	134	SSRERKRLVEALEKKELEKLETLINQVKTNILVE-----ELNRFR-----	175
Db	1750	AMKEVEKPVESAPEKDNESLEAPEINIEPIRVLVETKIMGPKSLNEDNDDDDGSECL	1809
Qy	176	-----RSSGVFRSSPLD--SLQDSITLSQQLFGSRD--GQTMGN-----PEQSLMDQ	220
Db	1810	DSIGDLSERTIQRFNTSIDPIRDSFSSISFGDRQKRFRTAENIRQDLPLPFQSSVQ	1869
Qy	221	IMRSAANPTV-----NPASLSPSLPPI---SDKEFQTK---BEDEQAD---EDEE	262
Db	1870	YLRSNPNPSQQLLVNLSMDSFSLSPNAPPVGFNTAQFLEKLQEDRPSAEGSIDSG	1929
Qy	263	MGQTHETKEAATAAEKNSKQSRVSTDSTQRPVAVSIGGDAVPVPSDDAGANCLGLDPVH	322
Db	1930	FEKVDHEGLDEFAAPPVH-----DPMQK---SVFGSLG-----SDD-----MKPGS	1967
Qy	323	QDGGPPIGHSFGLSAALDADRYLLESQLLASPNASTVDDVLAGD--SAACFTNPLPS-	379
Db	1968	QDGGFVIERNEANATLK-----KNQKSHHNDVIEKNYFNNDNAPTALESFIAEE	2021
Qy	380	-----DYDFDINDFLTD-----DANHAAYDIVAASN	405
Db	2022	ARKLVQDAVESAYKKQAVDSGEIGRELLDNVEQKIBQVKPEIVDSLHKAYDVGVDV	2081
Qy	406	YAAADRELDLEIHPENQIPSR---HSIQOQSGASSHGCD	444
Db	2082	HETVPNAVDDFVREAEKQLPESVPPEKIETPPLVDIHTVD	2123
RESULT 8			
S16706			
transcription activator YAP1 - yeast (Saccharomyces cerevisiae)			
N:Alternate names: hyper-resistance-conferring protein SNQ3; PAR1 protein; protein YM957			

C:Species: Saccharomyces cerevisiae			
C>Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 09-Jul-2004			
C:Accession: S16706; S19639; A30165; S48512; JE0416; S55112; S14832; S17002			
R:Hertle, K.; Haas, E.; Brendel, M.			
Curr. Genet. 19, 429-433, 1991			
A>Title: The SNQ3 gene of Saccharomyces cerevisiae confers hyper-resistance to several f			
A:Reference number: S16706; MUID:91347405; PMID:1878996			
A:Accession: S16706			
A:Molecule type: DNA			
A:Residues: 1-650 <HER>			
A:Cross-references: UNIPROT:P19880; UNIPARC:UPI0000052EC3; EMBL:X63268; NID:G4505; PIDN:			
R:Schnell, N.; Entian, K.D.			
Eur. J. Biochem. 200, 487-493, 1991			
A>Title: Identification and characterization of a Saccharomyces cerevisiae gene (PAR1) c			
A:Reference number: S19639; MUID:91364700; PMID:1889413			
A:Accession: S19639			
A>Status: nucleic acid sequence not shown; not compared with conceptual translation			
A:Molecule type: DNA			
A:Residues: 1-650 <SCH>			
A:Cross-references: UNIPARC:UPI0000052EC3			
R:Moyle-Rowley, W.S.; Harehman, K.D.; Parker, C.S.			
Genes Dev. 3, 283-292, 1989			
A>Title: Yeast YAP1 encodes a novel form of the jun family of transcriptional activator I			
A:Reference number: A30165; MUID:89252807; PMID:2542125			
A:Accession: A30165			
A:Molecule type: DNA			
A:Residues: 1-585 'E', 587-650 <MOY>			
A:Cross-references: UNIPARC:UPI0000168EFE; GB:X58693; NID:G4797; PIDN:CAA41536.1; PID:G4:			
R:Schnell, N.			
submitted to the EMBL Data Library, June 1991			
A:Reference number: S48512			
A:Accession: S48512			
A:Molecule type: DNA			
A:Residues: 1-315 'S', 317-650 <SCH2>			
A:Cross-references: UNIPARC:UPI0000069D64; EMBL:X60780; NID:G499192; PID:G499193			
R:Huesain, M.; Lenard, J.			
Gene 101, 149-152, 1991			
A>Title: Characterization of PDR4, a Saccharomyces cerevisiae gene that confers pleiotro			
A:Reference number: JE0416; MUID:91285426; PMID:2060792			
A:Accession: JE0416			
A:Molecule type: DNA			
A:Residues: 1-290, 'L'KLNLMNKFSLVR', 305-647, 'D', 649-650 <HUS>			
A:Cross-references: UNIPARC:UPI0000168D62; GB:X53830; NID:G4121; PIDN:CAA37827.1; PID:G4:			
R:Gentles, S.; Bowman, S.			
submitted to the EMBL Data Library, June 1995			
A:Reference number: S55102			
A:Accession: S55112			
A:Molecule type: DNA			
A:Residues: 1-650 <GEN>			
A:Cross-references: UNIPARC:UPI0000052EC3; EMBL:Z49810; NID:G854472; PID:G854483; MIPS:Y			
A:Experimental source: strain AB972			
C:Genetics:			
A:Gene: SGD:YAP1; PDR4; SNQ3; PAR1			
A:Cross-references: SGD:S0004466; MIPS:YML007W			
A:Map position: 13L			
C:Function:			
A>Description: transcription activator			
A>Note: confers pleiotropic drug resistance when present in high copy number			
C:Keywords: antibiotic resistance; DNA binding; leucine zipper; nucleus; transcription r			
F:59-99/Domain: fos/jun DNA-binding domain homology <FJD>			
F:63-156/Domain: DNA binding #status predicted <DNB>			
F:92-120/Region: leucine zipper motif			
Query Match			
Best Local Similarity 6.5%; Score 148.5; DB 2; Length 650;			
Matches 106; Conservative 85; Mismatches 171; Indels 239; Gaps 23;			
Qy	44	DMWTPDSVADIDSRLSVIPESQDAEDSHSATSATPSTSEKKPVKGRKSGWQVLPEPKT	103
Db	11	DVVSFGLAEFEG-----SKSEHDEIENEHRTGTGTGDSQPKKGS-----	54
Qy	104	NLPPRKRAKTEDEKQRRVERVLRNRAAQSSRRKRLVEALEKRNKELETLINQVKT	163


```
Db 55 -----KTSKKQDLDPETKQKRTAQNRAAQAFAFRKPKKKEKKVQSLE-----SIQQQ 105
QY 164 N-----LILVELNRFR-----RSSGVVTRSSPLDS 190
Db 106 NEVEATFLRDLQILTLVNLKLYKRPETRNDKSVLEYLARRDPLNLHFSKNVNHNSNPEIDT 165
QY 191 LODSI-----TSLQQLFGSRDGTMSNPEQSL-----MDQIMRSAAN 227
Db 166 PNDIOENVKQKNFTFQYPLDNDNDNSKNVGKQLPSPNDFSHAPINQITQKCLSD 225
QY 228 PTNPNASLSPSPISD-----KEFQTKKEEBEQ 257
Db 226 ATDSSSATLDSLNSNDVNLNTPNSSTMDLNDVLYTNRFVSGDDGSKNTKQLDSNMF 285
QY 258 DDEEMEQTWHE-----TKB-----AAAAKEKSKQSRVSTDSQTPA 295
Db 286 SDFNFENQFDEQSEFCSKMNQVCRQCPKPKPISALDKVEFASISLNSNS---PA 342
QY 296 VSIIGDAAVPFSDDAGANCLGDPVHODDGPFSIGHSGF--LSAALDADRYLLESQLLAS 354
Db 343 LNTWESHNI--TDNTPANVIATDATKYENS--FS---GFGRLGFDMSAHNVVNDNSSTGS 397
QY 355 PNA-----STVDDDDL-----AGDSAAACFTNP--- 376
Db 398 TDSGTGNGKNKNNNSDDVLPFISESPDMNQVNTNFPSPGSGTGIGNNAASNTNPSLLQ 457
QY 377 -----LPSD-----YDFDI-----NDFLTD 391
Db 458 SSKEDIPFINANLAFDDNSTNIQLQPPSESQSKNQKFDYDMFPRDSSKEGNNLPGEFLED 517
QY 392 DANHAAYDIVAASNYAAADREL---DLEIHDPE-----NQIPSRHSIQPOSGASHGCC 442
Db 518 DB-----DDKKAANSDDSESLIKQLINEPELPKQYLQSVPGNESEISQKXGSLQNA 572
QY 443 D 443
Db 573 D 573

RESULT 9
T15348
hypothetical protein B0350.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T15348
R:Gattung, S.
submitted to the EMBL Data Library, February 1996
A:Description: The sequence of C. elegans cosmid B0350.
A:Reference number: Z18332
A:Accession: T15348
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5170 <GAT>
A:CROSS-references: UNIPROT:Q17450; UNIPARC:UPI0000168675; EMBL:U50071; NID:g1208871; PI
C:Genetics:
A:Gene: CESP:B0350.1
A:Introns: 48/1; 5039/3; 5116/3

Query Match 6.5%; Score 148.5; DB 2; Length 5170;
Best Local Similarity 19.2%; Pred. No. 5.9;
Matches 96; Conservative 76; Mismatches 203; Indels 125; Gaps 19;

QY 16 PAESFLSAPGDNFTSLFADSTPST-----LNPRDMTTPDSVAIDRLSLVIP----- 62
Db 3757 PEDSHSPVPSDDVHGFKVTKITTTTTTTHHEPEPDHTSDEHVESERYASGSPVPSSEN 3816
QY 63 -----ESQDAEDESHSTSATAPSTSEKKPVYKRSWQVLPKTNLPP 107
Db 3817 SNRVTTTTTTTTHREPEPDQEHVVESQSYAS--GSPVPSKSVKVIETTTTTT 3875
QY 108 RKRKATEDEKQRRVRLNRRAAQSRKRKRLEVALEKKNKELE--TLLINVOKTNL 165
Db 3876 REHPHEDE-----IPTIVESHDDQAAS-----SVPSEEDVHGQIQITTTTTT 3925
```

```
QY 166 ILVBEINFRSSGVVTRSSPLDSLOQSITLSQ-----QLFGSRD 206
Db 3926 VPDEIDSGRMDELEKYSSESPVPEEDSSRVIEITTTTTTTFIREHPEPDDHSHVVGSOE 3985
QY 207 GQTMNSN--PEQSLMDQIMRSAANPTVNPAASLPSPILSDKEFQTKEE-----DEBQ 256
Db 3986 YSASGSPVPSKSVSRVIEITTTTTT-----TSEHFEHEDEHAQLHDTVSSBK 4033
QY 257 ADEDEMEQTHETKEAAAAKEKSKQSRVSTDSQTPAASVIGDAAVPFSDDAGANCL 316
Db 4034 TSSSEPVQTLLEAADIETHEQRFSAEDEHQPTSTKPV-----HCFMETTSTSHASHV 4088
QY 317 GLDPVHQDGPFSIGHSGF--LSAALDADRYLLESQLLASPNASTVDDYLAGSAAACF 373
Db 4089 ELE--HRDDDDSGEGGFGSKVLGPAKKAG--MVAGGVVAAPVALAA-----VGAKAA-- 4137
QY 374 TNPLPSDYDFDINDPLTDANHAAYDIVAASNYAAADRE---LDEIHDPEINQIPSRHS 429
Db 4138 -----YD-----ALKODDEATYDL-----DREHGSYESAELQRPQQVESROE 4177
QY 430 IQPOSGASHGCDGDIIV 449
Db 4178 DSESGSPVDHGDSQHLPV 4197

RESULT 10
A59267
myosin X - human
C:Species: Homo sapiens (man)
C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004
C:Accession: A59267
R:Rogers, M.S.; Strehler, E.B.
submitted to GenBank, March 2000
A:Description: Identification of myosin X as a specific binding partner for the tumor ser
A:Reference number: A59267
A:Accession: A59267
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-2058 <ROG>
A:CROSS-references: UNIPROT:Q9HD67; UNIPARC:UPI0000042862; GB:AF234532; PIDN:AAF37875.1
C:Genetics:
A:Gene: MYO10
F:66-727/Domain: myosin motor domain homology <MMO>

Query Match 6.4%; Score 147.5; DB 2; Length 2058;
Best Local Similarity 23.5%; Pred. No. 2.2;
Matches 86; Conservative 58; Mismatches 119; Indels 103; Gaps 18;

QY 91 RKSNGQVLPEPKTNLPKPK-----RAKTEDEKQRRVRLNRRAAQSRRKRKLEVEA 145
Db 805 RRVYQLLAKEQOEKKQBEKKKEBEERERERERERERERERERERERERERERERER 864
QY 146 LEKRNKELE--TLLINVOKTNLILVEELNRRFRSSGVVTRSSPLDSLOQSITLSQQLFGS 204
Db 865 LQKSQKEAELTRELEKQKKNK--QVEEILRLK-----EIEDLQRMKEQ----- 906
QY 205 RDGQTMNSPEQSLMDQIMRSAANPTVNPAASLPSPILSDKEFQTKKEEBEQ----- 257
Db 907 ---QELSITIASLQ-----KLQERRDQELRLLEEEACRAAQEFLES 944
QY 258 ---DEDER-----MEQTHETKEAAAAKEKSKQSRVSTDSQ---RPAVSTGGDAAPVF 307
Db 945 LNFIDEICVNRNIRSLSVGSEFSELAESAACEKPNFNPQPYPEEVEDEGFEADDAF 1004
QY 308 SDDAGANCLGDPVHODDGPFSIGHSGF--FGLSALDA---DRYLLSOLLASPN--ST 359
Db 1005 KD-----SPNPEHGHSDORTSGIRTSDDSEEDPYMNVDTVPTSPASST 1050
QY 360 V-----DDDL-----AGDSAAACF---TNPLPS---DYDFDINDPLTDANHAAYDIVA 402
Db 1051 VLLAPSVQDSSGLHNSSGESTYCMQPQWAGDLPDPGDYDYDQDDY--EDGAIITSGSVTF 1109
```

QY	403	ASNYAA	408
		::: :	
D _b	1110	SNSYGS	1115

RESULT 11

JC6009
 surface-located membrane protein lmp3 precursor - Mycoplasma hominis
 C/Species: Mycoplasma hominis
 C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
 C/Accession: JC6009
 R/Ladefoged, S.A.; Jensen, L.T.; Brock, B.; Birkelund, S.; Christiansen, G.
 J. Bacteriol. 178, 2775-2784, 1996
 A/Title: Analysis of 0.5-kilobase-pair repeats in the Mycoplasma hominis lmp gene system
 A/Reference number: JC6009; MUID:96213016; PMID:8631654

A:Accession number: JC6009
A:Molecule type: DNA
A:Residues: 1-1302 <LAD>
A:Cross-references: UNIPROT:Q49547; UNIPARC:UIP0000B0501; EMBL:X95601; NID:g1197335; PIR:P08003
C:Genetics:
A:Gene: lmp3
A:Genetic code: SGC3
C:Superfamily: surface-located membrane protein lmp3; tetratricopeptide repeat homology
C:Keywords: duplication; membrane protein
F:1-24/Domain: signal sequence #status Predicted <SIG>
F:25-1302/Product: surface-located membrane protein Lmp3 #status predicted <MAT>
F:937-992/Domain: tetratricopeptide repeat homology <Tt1>
F:993-1026/Domain: tetratricopeptide repeat homology <Tt2>
F:1089-1120/Domain: tetratricopeptide repeat homology <TT3>
F:1154-1190/Domain: tetratricopeptide repeat homology <TT4>

Query Match	6.4%	Score 146;	DB 1;	Length 1302;
Best Local Similarity	20.7%	Pred. No. 1.5;		
Matches 97;	Conservative	75;	Mismatches	200;
Indels	96;	Gaps	18;	

[illegible]

```

QY 140 RLEV-EALEKN-----KELETL-----LINQKTN---LILVELNFRSSGVVT 182
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 861 VTEITKLETFENKDKVFKLEQTRKQIDEIFNTKNTNDPYSTLI BELTSKRDSKNSIT 920
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 183 RSSPSLQSDISITLSQOLFSGRDQTNMSPNQSLMDQIMRSA---ANPTVNPASLASPSL 239
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 921 NSSNKSQDETANTELKQAL-----AKANTDKQADNIAPSTKEQIANKSI--SSANTLL 971
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

[illegible]

QY	335	GLSAALDADRYLLESOLIASPNASTVDDDYLAGDSAACTNPLPSDYDFDINDF---	LTD	391
	:	:	:	:
DB	1081	NKSDIETANTELT--KOALA KANTDYAQADNLARSTKEQLNKSISS-----	ANTLLAKLTD	1133
	:	:	:	:
QY	392	DANHAAYDIVAASNVAAADRELDLEIHDPENQIIPSRHSIQOQSGASS		439
	:	:	:	:

```

RESULT 12
T18519
myosin X - bovine
C.Species: Bos primigenius taurus (cattle)
C.Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text_change 09-Jul-2004

```

C;Accession: T18519
R;Corey, D.P.; Deffler, B.H.; Solc, C.K.; Duyk, G.M.; Cheney, R.E.
submitted to the EMBL Data Library, April 1996
A;Description: Cloning and expression of myosin X, a novel unconventional myosin with ple
A;Reference number: Z18942

A:Accession: I18519
A:Status: Preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-2052 <OR>
A:Cross-references: UNIPROT:P79114; UNIPARC:UPI000012FABC; EMBL:U55042; NID:gl755048; PII
A:Experimental source: aorta
C:Keywords: nucleotide binding; P-loop
F:66-727/Domain: myosin motor domain homology <MMO>
F:157-164/Region: nucleotide-binding motif A (P-loop)

Query Match	6.3%	Score 145;	DB 2;	Length 2052;
Best Local Similarity	24.1%	Pred. NO. 3;		
Matches 82;	Conservative 43;	Mismatches 119;	Indels 96;	Gaps 15;

[illegible]

```

199 QY QQLFGSRDQGTMSNPQSLMDQIMRSAANTVNPASLSPSLPPI SDKERQTKEDEQA- 257
      : : : : :
903 ---RMKREQSLSTEASLQ-----KLOQLRDEELRLLEDEACRAA 939
      : : : : :
258 -----DEDEEMEQTWHE-----TKAAAAKKNQSKRVSTSTDRPAVSIGG 300
      : : : : :
940 QY QYQYESTNEDTDEGTNITPESL SVGSCCTCEQGAKEPS---ENFQSDVPREEYDEGF 997
      : : : : :

```

```

301 DAAVPFVSDGAGANCLGLDPVHQDDPFSIGHSGFLSAAALDADRYLLESQILASPNA-ST 359
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
998 EADDDAFKDSNPSPSEHG---HSDQRTSGIRTS---DESSEEDPYMNDTVVETSPSADST 1050
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
360 V-----DDYLAGSAACTFN-----PLP-SPDYDFDINDF 388
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 13
SF0252
FUN12 protein - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: protein YAL0035w
C:Species: *Saccharomyces cerevisiae*

Accession: S70292; S51984; S48519
R.V.O., D.
submitted to the EMBL Data Library, April 1996
A:Reference number: S70291
A:Accession: S70292
A:Molecule type: DNA
A:Residues: 1-1002 <VOD>

[illegible]

LMSR, 472-640, 'Q', 642-721, 'H', 723-969, 'T', 971-1002 <805>
 A.Cross-References: UNIPARC:U1000017B294; EMBL:U12980; MIPS:YAL035w
 A.Note: This sequence has been revised in reference S70291
 R.Stratve, P.; Shater, B.K.; Strathern, J.N.; Hughes, S.H.
 submitted to the EMBL Data Library, March 1994
 A.Description: Isolation, identification and characterization of the FUN12 gene of *Saccharomyces cerevisiae*.
 A.Reference number: S48519
 S48510

A:Molecule type: DNA
A:Residues: 'MLK', 134, 'LLLRLLLRPNKL', 149, 'VKVINQVILRSQPR', 165, 'YCRGCEFTS', 177-180, 'NLSMR', 472-640, 'Q', 642-721, 'H', 723-965, 'I', 971-1000, 'SNEHGIR' <SUT>
A:Cross-references: UNIPARC:UPI0000168C1B; EMBL:L29389; NID:G460043; PIDN:AAAS7228.1; PIR:Q00000
C:Genetics:
A:Gene: SGD:FUN12
A:Cross-references: SGD:S0000033; MIPS:YAL035W
A:Map position: 1L
C:Keywords: GTP binding; nucleotide binding; P-loop; transmembrane protein
F:406-533/Domain: translation elongation factor Tu homology <ETU>
F:412-419/Region: nucleotide-binding motif A (P-loop)
F:491-507/Domain: transmembrane #status predicted <TM1>
F:530-533/Region: GTP-binding NKXD motif

Query Match 6.3%; Score 144; DB 2; Length 1002;
Best Local Similarity 24.0%; Pred. No. 1.4;
Matches 60; Conservative 41; Mismatches 69; Indels 80; Gaps 10;

QY 62 PESQDAEDDESHSTSATAPST-----SEKPKVKRKSQGVLPPEKTNLPPR 108
DB 35 PESSAGADTTSREASASAGAEIEGDFMSTLUKQKKQKQK-----VIEKKQKGPIL 88
QY 109 KRAKTEDEKEQRRVERVLNRRAAQSSRRERKLEVEALEKRNKELETLLINVOKTNLILV 168
DB 89 K-SKKEKEKEKEKQKKKEQAA-----RKAQQAQKQKKNEL-----NKQVKEKAAA 137
QY 169 EE--LNRFRSSGVVTRSSSP-----LDSLODSITISQQLFGSRDQGTWSNPPQ 215
DB 138 EKAAAEKSKQSGESDKPSASAKPAKVPAGLAALRRQLKQKL-----EQ 186
QY 216 SLMDQIMRSAAANTVNPASLSPLSPIDKEFTKREDEEQADEMEQWHTKEAAA 275
DB 187 EKLER-----BEEERLEKEBERLANEKKMK-----EKAARK 220
QY 276 AKEKNSQSR 285
DB 221 EKEKAKREKR 230

RESULT 14
TS0922
bZIP protein HV5 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004
A:Accession: TS0922
R:Oyama, T.; Shimura, Y.; Okada, K.
A:Description: The Arabidopsis HV5 gene encodes a bZIP protein that regulates stimulus-i
A:Reference number: Z25271
A:Accession: TS0922
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-168 <OYA>
A:Cross-references: UNIPROT:Q24646; UNIPARC:UPI000012CF4; EMBL:AB005295; PIDN:BA21116
A:Experimental source: Landsberg erecta

Query Match 6.3%; Score 143.5; DB 2; Length 168;
Best Local Similarity 30.5%; Pred. No. 0.18;
Matches 47; Conservative 19; Mismatches 57; Indels 31; Gaps 4;

QY 29 TSLFADSTPTLNPRDMTP-----DSVADIDSRLSVIP-----ESQDAEDDESHSTSATAP 80
DB 7 SSLAASSLPSSERSSSSAPHLEIKEGIESDEIRRVPEFGGEAVGKTSRGSGSAT-- 64
QY 81 STSEKKPKVKRKSQGVLPPEKTNLPPRKAETDEKEQRRVERVLNRRAAQSSRRERK 140
DB 65 -----CQRTQATVGESQRRKGTTPAEKKNKLLLRNRVSAQQAERKK 110
QY 141 L-----EVEALEKRNKELETLLINVOKTNLIL 167
DB 111 AYLSELENRVKOLENKENSELERLSTLQENQML 144

RESULT 15

T08591
TGACG-motif binding protein STF1 - soybean
C:Species: Glycine max (soybean)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 31-Dec-2004
A:Accession: T08591
R:Cheong, Y.H.; Yoo, C.M.; Park, J.M.; Ryu, G.R.; Goeckjian, V.H.; Nagao, R.T.; Key, J.L.;
submitted to the EMBL Data Library, September 1995
A:Description: STF1 is a novel TGACG-binding factor with a zinc-finger motif and a bZIP
A:Reference number: Z16445
A:Accession: T08591
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-326 <CHE>
A:Cross-references: UNIPROT:Q39895; UNIPARC:UPI00000A6233; EMBL:L28003; NID:G2934883; PIR
A:Experimental source: strain Williams; hypocotyl
C:Superfamily: TGACG-motif-binding transcription factor

Query Match 6.3%; Score 143.5; DB 2; Length 326;
Best Local Similarity 29.1%; Pred. No. 0.4;
Matches 43; Conservative 27; Mismatches 51; Indels 27; Gaps 5;

QY 48 PDSVADIDSRLSVIPESQDAEDDESHSTSATAPSTSEKKPKVKRKSQGVLPPEKTNLPP 107
DB 195 PES-----DEEIRRVPEI-----GGESAGTSASRPDAGSNAGTERAQTGD-----SQ 237
QY 108 RKRKTEDEKEQRRVERVLNRRAAQSSRRERK-----LEVEALEKRNKELETLLINV 160
DB 238 KKRGRSPADKESKRLKLLNRVSAQAARERKKAYLIDLETTRVKOLEKKNSELKERLSTL 297
QY 161 QKTNLIL-----VEELNRFRRSSGVVTRSS 185
DB 298 QNENQMLRQILKNTTASRRSGNSGTNNA 325

Search completed: November 23, 2005, 03:26:02
Job time : 34.6947 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:04:00 ; Search time 176.139 Seconds
(without alignments)
1125.017 Million cell updates/sec

Title: US-10-663-450-2
Perfect score: 2294
Sequence: 1 MAFQSSPLVKFEAPSPSF.....QPQSGASHGCGDDGGIANGV 451

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*
9: Geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2294	100.0	451	5 AAE15371	Aae15371 Trichoder
2	2294	100.0	451	8 ADS12801	Ads12801 Trichoder
3	2294	100.0	451	8 ADS12804	Ads12804 Trichoder
4	2279.5	99.4	450	4 AAB82975	Aab82975 Trichoder
5	437.5	19.1	350	4 AAB82976	Aab82976 Aspergill
6	437.5	19.1	350	5 AAE15372	Aae15372 Aspergill
7	434.5	18.9	386	5 AAE15379	Aae15379 Aspergill
8	434.5	18.9	386	8 ADS12815	Ads12815 Aspergill
9	434	18.9	349	8 ADS12805	Ads12805 Aspergill
10	434	18.9	349	8 ADS12803	Ads12803 Aspergill
11	434	18.9	349	8 ADS12863	Ads12863 Aspergill
12	429.5	18.7	409	7 ABO43144	Abo43144 A. thalia
13	429.5	18.7	409	7 ADS13925	Ads13925 Plant (A.
14	429.5	18.7	409	8 ADO02271	Ado02271 Thalecres
15	426.5	18.6	342	4 AAB82977	Aab82977 Aspergill
16	426.5	18.6	342	5 AAE15381	Aae15381 Aspergill
17	426.5	18.6	342	8 ADS12818	Ads12818 Aspergill
18	343	14.1	64	5 AAE15373	Aae15373 Trichoder
19	288	12.6	64	5 AAE15374	Aae15374 Aspergill
20	242.5	10.6	230	2 AAW53806	Aaw53806 Transcrip
21	242.5	10.6	230	8 ADT87049	Adt87049 Yeast Str
22	242.5	10.6	238	2 AAW53807	Aaw53807 Transcrip
23	231.5	10.1	200	8 ADS43437	Ads43437 Bacterial
24	196	8.5	68	5 AAE15382	Aae15382 Yeast HAC

25	196	8.5	68	8	ADS12859	Saccharom
26	196	8.5	84	5	ABP02534	Human ORF
27	166	7.2	3257	4	ABB67502	Abb67502 Drosophi
28	165.5	7.2	192	4	AAB82615	Aab82615 Maize tra
29	165.5	7.2	192	4	AAB82616	Aab82616 Maize roo
30	165.5	7.2	192	4	AAG66526	Aag66526 Maize roo
31	163.5	7.1	1442	4	ABG25601	Abg25601 Novel hum
32	163.5	7.1	2408	4	ABG10631	Abg10631 Novel hum
33	163	7.1	192	4	AAB82614	Aab82614 Maize roo
34	163	7.1	192	4	AAG66525	Aag66525 Maize roo
35	148.5	6.5	650	6	ABR52838	AbR52838 Plectu
36	148.5	6.5	650	7	ADK62274	Adk62274 Drosasa
37	148.5	6.5	869	3	AAB42020	Aab42020 Human ORF
38	148.5	6.5	2048	4	AAE11891	Aae11891 Angiogene
39	148.5	6.5	2057	4	AAE11890	Aae11890 Angiogene
40	147.5	6.4	2058	5	ABB97219	Abb97219 Novel hum
41	147.5	6.4	2058	6	ABP71989	Abp71989 Human myo
42	147.5	6.4	2058	6	ABU03558	Abu03558 Angiogene
43	147.5	6.4	2058	7	ABO84401	Abo84401 Human myo
44	147.5	6.4	2058	8	ADL61264	Adl61264 Human pro
45	147.5	6.4	2058	8	ADQ17925	Adq17925 Human sof

ALIGNMENTS

RESULT 1
AAE15371
ID AAE15371 standard; protein; 451 AA.

XX AAE15371;

XX
DT 29-AUG-2003 (revised)
DT 07-MAR-2002 (first entry)

XX Trichoderma reesei HAC1 protein.

XX Heterologous protein secretion; unfolded protein response; UPB; lipase;
KW cellulase; carbohydrase; industry; purification; HAC1 protein.

XX Hypocrea jecorina.

XX Key Location/Qualifiers

FT Binding-site 84..147
FT /label= DNA-binding_domain

XX US2001034045-A1.

XX 25-OCT-2001.

XX 23-MAR-2001; 2001US-00816277.

XX 24-MAR-2000; 2000US-00534692.

XX (GENV) GENENCOR INT INC.

XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MJA;

XX WPI: 2002-033728/04.

XX N-PSDB; AAD24595.

XX Increasing secretion of heterologous proteins e.g. lipase and cellulase
in eukaryotic cells useful in industry to increase production and
facilitate purification, by inducing an elevated unfolded protein
response.

XX Claim 39; Fig 7; 56pp; English.

XX The present invention relates to methods for increasing the secretion of
heterologous protein in eukaryotic cells by inducing an elevated unfolded
protein response (UPR). The method involves inducing the elevated UPR by
increasing the presence of proteins such as HAC1, HACA, PTC2 or PRF1 in
cells. The method and sequences are useful for increasing the secretion

CC of heterologous proteins (e.g. lipase, cellulase, carbohydrase) in
CC eukaryotic cells useful in industry to increase protein yields and to
CC facilitate purification. The present sequence is Trichoderma reesei HAC1
CC protein. (Updated on 29-AUG-2003 to standardise OS field)
XX
SQ Sequence 451 AA;

Query Match 100.0%; Score 2294; DB 5; Length 451;
Best Local Similarity 100.0%; Pred. No. 3.7e-179;
Matches 451; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAFQSSPLVKFEASPAESFLSAPGDNFTSLPADSTPTLNPRDMTPDSVADIDSRLSV 60
DB 1 MAFQSSPLVKFEASPAESFLSAPGDNFTSLPADSTPTLNPRDMTPDSVADIDSRLSV 60

QY 61 IPESQDAEDDESHSTSATAPSTSEKKPVKKRSGQVLPEPKTNLPKRKTEDEKEOR 120
DB 61 IPESQDAEDDESHSTSATAPSTSEKKPVKKRSGQVLPEPKTNLPKRKTEDEKEOR 120

QY 121 RVERVLNRRAAQSSRRERKLEVEALEKRNKELETLINNVQKTNLILVEELNFRSSGV 180
DB 121 RVERVLNRRAAQSSRRERKLEVEALEKRNKELETLINNVQKTNLILVEELNFRSSGV 180

QY 181 VTRSSPLDSLQDSITLSQQLFGSRDGTMSNPEQSLMDQIMRSAANPTVNPASLSPSLP 240
DB 181 VTRSSPLDSLQDSITLSQQLFGSRDGTMSNPEQSLMDQIMRSAANPTVNPASLSPSLP 240

QY 241 PISDKFEQTKEDDEQADEMEQTHETKAAAKENKSKOSRVSTDTORPAVSI 300
DB 241 PISDKFEQTKEDDEQADEMEQTHETKAAAKENKSKOSRVSTDTORPAVSI 300

QY 301 DAAVPVFSDDAGANCLGLDPVHQDDGPFSGHSGFLSAAALDADRYLLESQLLASPNASTV 360
DB 301 DAAVPVFSDDAGANCLGLDPVHQDDGPFSGHSGFLSAAALDADRYLLESQLLASPNASTV 360

QY 361 DDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYYAAADRELDLEIHPD 420
DB 361 DDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYYAAADRELDLEIHPD 420

QY 421 ENQIPSRHSIQPQSGASSHGCGDGGIAGV 451
DB 421 ENQIPSRHSIQPQSGASSHGCGDGGIAGV 451

RESULT 2
ADS12801
ID ADS12801 standard; protein; 451 AA.
AC ADS12801;
XX
XX 16-DEC-2004 (first entry)
XX Trichoderma reesei hac1 chaperone and foldase.
DE
XX unfolded protein response; UPR; HAC1; PTC2;
KW unfolded protein response modulator; enzyme production; hac1; chaperone;
KW foldase; enzyme.
XX
OS Hypocrea jecorina.
XX
PN US2004186070-A1.
XX
XX 23-SEP-2004.
PD
XX 15-SEP-2003; 2003US-00663450.
PF
XX 24-MAR-2000; 2000US-00534692.
PR 23-MAR-2001; 2001US-00816277.
XX
XX (GENV) GENENCOR INT INC.
PA
XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MJA;
PI
XX

DR WPI; 2004-707924/69.
DR N-PSDB; ADS12800.
XX
PT Increasing the secretion of a heterologous protein, such as a therapeutic
PT or an industrial enzyme, in genetically modified eukaryotic cells by
PT inducing an elevated unfolded protein response (UPR).
XX
PS Example 3; SEQ ID NO 2; 83pp; English.
XX
CC The invention describes a method of increasing the secretion of a
CC heterologous protein in a eukaryotic cell, comprising inducing an
CC elevated unfolded protein response (UPR). Also described are: an isolated
CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
CC unfolded protein response and has less than 50% similarity to yeast HAC1
CC protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
CC protein induces unfolded protein response and where the HAC1 protein
CC comprises a DNA binding region that has greater than 70% similarity to
CC the DNA binding region of filamentous fungi HAC1 protein; a protein
CC having unfolded protein response inducing activity and having greater
CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
CC respectively) as given in the specification; a protein having an amino
CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
CC response and has at least 70% similarity to a fully defined amino acid
CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
CC specification; an isolated nucleic acid encoding a IRE1 protein that
CC modulates unfolded protein response and has at least 60% similarity to a
CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
CC having unfolded protein response modulating activity and having greater
CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
CC heterologous nucleic acid encoding a protein having unfolded protein
CC response modulating activity and a heterologous nucleic acid encoding a
CC protein of interest to be secreted. The methods and compositions of
CC genetically manipulating cells to have an elevated unfolded protein
CC response (UPR) resulting in an increased capacity to produce secreted
CC proteins, are useful in e.g. production of therapeutic or industrial
CC enzymes. This is the amino acid sequence of Trichoderma reesei HAC1
CC chaperone and foldase.
XX
SQ Sequence 451 AA;

Query Match 100.0%; Score 2294; DB 8; Length 451;
Best Local Similarity 100.0%; Pred. No. 3.7e-179;
Matches 451; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAFQSSPLVKFEASPAESFLSAPGDNFTSLPADSTPTLNPRDMTPDSVADIDSRLSV 60
DB 1 MAFQSSPLVKFEASPAESFLSAPGDNFTSLPADSTPTLNPRDMTPDSVADIDSRLSV 60

QY 61 IPESQDAEDDESHSTSATAPSTSEKKPVKKRSGQVLPEPKTNLPKRKTEDEKEOR 120
DB 61 IPESQDAEDDESHSTSATAPSTSEKKPVKKRSGQVLPEPKTNLPKRKTEDEKEOR 120

QY 121 RVERVLNRRAAQSSRRERKLEVEALEKRNKELETLINNVQKTNLILVEELNFRSSGV 180
DB 121 RVERVLNRRAAQSSRRERKLEVEALEKRNKELETLINNVQKTNLILVEELNFRSSGV 180

QY 181 VTRSSPLDSLQDSITLSQQLFGSRDGTMSNPEQSLMDQIMRSAANPTVNPASLSPSLP 240
DB 181 VTRSSPLDSLQDSITLSQQLFGSRDGTMSNPEQSLMDQIMRSAANPTVNPASLSPSLP 240

QY 241 PISDKFEQTKEDDEQADEMEQTHETKAAAKENKSKOSRVSTDTORPAVSI 300
DB 241 PISDKFEQTKEDDEQADEMEQTHETKAAAKENKSKOSRVSTDTORPAVSI 300

QY 301 DAAVPVFSDDAGANCLGLDPVHQDDGPFSGHSGFLSAAALDADRYLLESQLLASPNASTV 360
DB 301 DAAVPVFSDDAGANCLGLDPVHQDDGPFSGHSGFLSAAALDADRYLLESQLLASPNASTV 360

QY 361 DDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYYAAADRELDLEIHPD 420
DB 361 DDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYYAAADRELDLEIHPD 420

QY 421 ENQIPSRHSIQPOSGASSHGCDGGIAGV 451
 Db 421 ENQIPSRHSIQPOSGASSHGCDGGIAGV 451

RESULT 3

ADSI2804
 ID ADSI2804 standard; protein; 451 AA.

AC ADSI2804;

DT 16-DEC-2004 (first entry)

XX Trichoderma reesei hac1 DNA binding domain.

DE unfolded protein response; UPR; HAC1; PTC2;

XX unfolded protein response modulator; enzyme production; hac1; chapertone;

KW foldase; DNA binding domain.

XX Hypocrea jecorina.

OS US2004186070-A1.

PN 23-SEP-2004.

XX 15-SEP-2003; 2003US-00663450.

PF 24-MAR-2000; 2000US-00534692.

PR 23-MAR-2001; 2001US-00816277.

XX (GEMV) GENENCOR INT INC.

PA Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;

XX WPI; 2004-707924/69.

DR Increasing the secretion of a heterologous protein, such as a therapeutic
 PT or an industrial enzyme, in genetically modified eukaryotic cells by
 PT inducing an elevated unfolded protein response (UPR).

XX Example 3; SEQ ID NO 5; 83pp; English.

XX The invention describes a method of increasing the secretion of a
 CC heterologous protein in a eukaryotic cell, comprising inducing an
 CC elevated unfolded protein response (UPR). Also described are: an isolated
 CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
 CC unfolded protein response and has less than 50% similarity to yeast HAC1
 CC protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
 CC protein induces unfolded protein response and where the HAC1 protein
 CC comprises a DNA binding region that has greater than 70% similarity to
 CC the DNA binding region of filamentous fungi HAC1 protein; a protein
 CC having unfolded protein response inducing activity and having greater
 CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
 CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
 CC respectively) as given in the specification; a protein having an amino
 CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
 CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
 CC response and has at least 70% similarity to a fully defined amino acid
 CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
 CC specification; an isolated nucleic acid encoding a IRE1 protein that
 CC modulates unfolded protein response and has at least 60% similarity to a
 CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
 CC having unfolded protein response modulating activity and having greater
 CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
 CC heterologous nucleic acid encoding a protein having unfolded protein
 CC response modulating activity and a heterologous nucleic acid encoding a
 CC protein of interest to be secreted. The methods and compositions of
 CC genetically manipulating cells to have an elevated unfolded protein
 CC response (UPR) resulting in an increased capacity to produce secreted
 CC proteins, are useful in e.g. production of therapeutic or industrial
 CC enzymes. This is the amino acid sequence of Trichoderma reesei HAC1
 CC chapertone and foldase DNA binding domain.

XX SQ Sequence 451 AA;

Query Match 100.0%; Score 2294; DB 8; Length 451;
 Best Local Similarity 100.0%; Pred. No. 3.7e-179;
 Matches 451; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAFQSSPLVKFEASPAESFLSAPGDNFTSLFADSTPSTLNPRDMTPDSVADIDSRLSV 60
 Db 1 MAFQSSPLVKFEASPAESFLSAPGDNFTSLFADSTPSTLNPRDMTPDSVADIDSRLSV 60
 QY 61 IPESQDAEDDESHSTSATAPSTSEKKPVKKRKSQGVLPKPTNLNPPRKRAKTEDEKEOR 120
 Db 61 IPESQDAEDDESHSTSATAPSTSEKKPVKKRKSQGVLPKPTNLNPPRKRAKTEDEKEOR 120
 QY 121 RVERVLNRRAAQSSRRERKRLEVEALEKRNKELETLLINVQKTNLLI LVEELNFRSSGV 180
 Db 121 RVERVLNRRAAQSSRRERKRLEVEALEKRNKELETLLINVQKTNLLI LVEELNFRSSGV 180
 QY 181 VTRSSSPLDSLODSITLSQOLFCSRDCGOTMSNPEQSLMDQIMRSAANPTVNPASISPLP 240
 Db 181 VTRSSSPLDSLODSITLSQOLFCSRDCGOTMSNPEQSLMDQIMRSAANPTVNPASISPLP 240
 QY 241 PISDKFEQTKEDDEQADEMEQTHETKEAAAAKEKNSKOSRVSTDTORPAVSIGG 300
 Db 241 PISDKFEQTKEDDEQADEMEQTHETKEAAAAKEKNSKOSRVSTDTORPAVSIGG 300
 QY 301 DAAVPVFSDDAGANCLGLDPVHODDGPFSIGHSGLSAALDADRYLLESOLLASPNASTV 360
 Db 301 DAAVPVFSDDAGANCLGLDPVHODDGPFSIGHSGLSAALDADRYLLESOLLASPNASTV 360
 QY 361 DDYLAGDSAACTNPDPDYDFDINDFLTDDANHAAYDIIVAASNYAAAHNENLTHLP 420
 Db 361 DDYLAGDSAACTNPDPDYDFDINDFLTDDANHAAYDIIVAASNYAAAHNENLTHLP 420
 QY 421 ENQIPSRHSIQPOSGASSHGCDGGIAGV 451
 Db 421 ENQIPSRHSIQPOSGASSHGCDGGIAGV 451

RESULT 4

AAB82975

ID AAB82975 standard; protein; 450 AA.

XX AAB82975;

AC AAB82975;

DT 11-SEP-2003 (revised)

DT 21-DEC-2001 (first entry)

XX Trichoderma reesei HAC1, involved in unfolded protein response

DE HAC1; transcription factor; unfolded protein response; protein secretion.

XX Hypocrea jecorina.

XX Key Location/Qualifiers

FT Domain 84..147

FT /label= DNA binding domain

XX WO200172783-A2.

XX 04-OCT-2001.

XX 23-MAR-2001; 2001WO-US009401.

XX 24-MAR-2000; 2000US-00534692.

XX (GEMV) GENENCOR INT INC.

XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo ML

XX WPI; 2001-626252/72.

XX N-PSDB; AAB26931.

XX Increasing secretion of heterologous proteins e.g. lipase and cellulase
PT in eukaryotic cells useful in industry to increase production and
PT facilitate purification, by inducing an elevated unfolded protein
PT response.

XX Claim 54; Fig 7A-B; 89pp; English.

XX The present sequence is that of the HAC1 protein of *Trichoderma reesei*,
CC as deduced from the newly isolated HAC1 gene (see AAH26931). HAC1 protein
CC is a transcription factor involved in the unfolded protein response
CC (UPR). The invention provides methods for increasing the secretion of a
CC heterologous protein in a cell by inducing an elevated UPR. This can be
CC achieved by modulating the activity of HAC1, PTC2 or IRE1 in the cell,
CC e.g. by gene overexpression. The cell from which the protein is secreted
CC can be any cell having an UPR, such as mammalian cells, insect cells,
CC yeast and filamentous fungi. The protein of interest can be any secreted
CC protein such as a therapeutic protein or an industrial enzyme, e.g.
CC lipase, cellulase, endoglucanase-H, protease, carboxylate, reductase,
CC oxidase, isomerase, transferase, kinase, phosphatase, alpha-amylase,
CC glucosylase, lignocellulose hemicellulase, pectinase and ligninase
CC (claimed). (Updated on 11-SEP-2003 to standardise OS field)

XX Sequence 450 AA;

Query Match 99.4%; Score 2279.5; DB 4; Length 450;
Best Local Similarity 99.8%; Pred. No. 5.7e-178;
Matches 450; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 MAFQSSPLVKFEASPAESFLSAGDNFTSLFADSTPTLNPRDMTPDSVADIDSRLSV 60

DB 1 MAFQSSPLVKFEASPAESFLSAGDNFTSLFADSTPTLNPRDMTPDSVADIDSRLSV 60

QY 61 IPESQDAEDESHSTATAPSTSEKKPVKRSWGQVLPPEKTNLPPRKRAKTEDEKEOR 120

DB 61 IPESQDAEDESHSTATAPSTSEKKPVKRSWGQVLPPEKTNLPPRKRAKTEDEKEOR 120

QY 121 RVERVLNRRAAQSSRRERKLEVEALEKRNKELETLLINVQKTNLLVLELNFRSSGV 180

DB 121 RVERVLNRRAAQSSRRERKLEVEALEKRNKELETLLINVQKTNLLVLELNFRSSGV 180

QY 181 VTRSSSPLSLQDSITLSQOLFSGRDGQTMNSPEQSLMDQIMRSAANPTVNPAISPSLP 240

DB 181 VTRSSSPLSLQDSITLSQOLFSGRDGQTMNSPEQSLMDQIMRSAANPTVNPAISPSLP 240

QY 241 PISDKFEQTKEDDEQADEMEQWTHETKAAAKKESKOSRVSTSTORPAVSGIG 300

DB 241 PISDKFEQTKEDDEQADEMEQWTHETKAAAKKESKOSRVSTSTORPAVSGIG 300

QY 301 DAAVPVFSDDAGANCLGLDPVHQDDGPFSGHSGFLSAAALDADRYLLESQLLASPNASTV 360

DB 301 DAAVPVFSDDAGANCLGLDPVHQDDGPFSGHSGFLSAAALDADRYLLESQLLASPNASTV 360

QY 361 DDDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAADRELIDLEIHP 420

DB 361 DDDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAADRELIDLEIHP 420

QY 421 ENQIPSRHSITQOPQSGASSHGCGDGGIAGV 451

DB 421 ENQIPSRHSITQOPQSGASSHGCGDGGIAGV 450

RESULT 5

AAB82976

ID AAB82976 standard; protein; 350 AA.

XX AAB82976;

XX 11-SEP-2003 (revised)

DT 21-DEC-2001 (first entry)

XX *Aspergillus nidulans* hacA, involved in unfolded protein response.

KW hacA; transcription factor; unfolded protein response; protein secretion.
XX *Emericella nidulans*.

XX Key Location/Qualifiers

FT Domain 53..116

FT /label= DNA binding domain

XX WO200172783-A2.

XX 04-OCT-2001.

XX 23-MAR-2001; 2001WO-US009401.

XX 24-MAR-2000; 2000US-00534692.

XX (GEMV) GENENCOR INT INC.

XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saikio, J, MJA,

XX WPI; 2001-626252/72.

XX N-PSDB; AAH26932.

XX Increasing secretion of heterologous proteins e.g. lipase and cellulase
PT in eukaryotic cells useful in industry to increase production and
PT facilitate purification, by inducing an elevated unfolded protein
PT response.

XX Claim 54; Fig 8A-B; 89pp; English.

XX The present sequence is that of the hacA protein of *Aspergillus nidulans*,
CC as deduced from the newly isolated hacA gene (see AAH26932). HacA protein
CC is a transcription factor involved in the unfolded protein response
CC (UPR). The invention provides methods for increasing the secretion of a
CC heterologous protein in a cell by inducing an elevated UPR. This can be
CC achieved by modulating the activity of HAC1 (or hacA), PTC2 or IRE1 in
CC the cell, e.g. by gene overexpression. The cell from which the protein is
CC secreted can be any cell having an UPR, such as mammalian cells, insect
CC cells, yeast and filamentous fungi. The protein of interest can be any
CC secreted protein such as a therapeutic protein or an industrial enzyme,
CC e.g. lipase, cellulase, endoglucanase-H, protease, carboxylate, reductase,
CC reductase, oxidase, isomerase, transferase, kinase, phosphatase, alpha-
CC amylase, glucosylase, lignocellulose hemicellulase, pectinase and
CC ligninase (claimed). (Updated on 11-SEP-2003 to standardise OS field)

XX Sequence 350 AA;

Query Match 19.1%; Score 437.5; DB 4; Length 350;
Best Local Similarity 32.4%; Pred. No. 4e-27;
Matches 146; Conservative 51; Mismatches 122; Indels 131; Gaps 18;

QY 26 DNFTSL-----FADSTPTLNPRDMTPDSVADIDSRLSVIPESQDAEDESHSTATAP 80

DB 5 DRFPVKMEDAFANSLPTTPPSLEVPLTVSPADTSLRTKNV-----VAOTKP 51

QY 81 STSEKKPVKRSWGQVLPPEKTNLPPRKRAKTEDEKEORRVERVLNRRAAQSSRRERK 140

DB 52 --EKKPAKRSWGQELPVPKTNLPPRKRAKTEDEKEORRVERVLNRRAAQSSRRERK 104

QY 141 LEVEALEKRNKELETLLINVQKTNLLVLELNFRSSGVVTRSSSPIDSLIDSLTTC 200

DB 110 LEMEKLESEK-----IDMEQNOQFLQRLQAQEAENNRLS--..OOVAQLSAF 141

QY 201 LFSGRDGQTMNSPEQSLMDQIMRSAANPTVNPAISPSLPISDKFEQTKEDDEQADE 450

DB 154 VRGSRH-----STPTSS-----SPASVSTLTPTLTKO-----EGDEVPLDRI 191

QY 260 --DEMEQWTHETKAAAKKESKOSRVSTSTORPAVSGIG-----DAAVPVFSDDAGA 313

DB 192 PPTPSVTVDYSPTLKPSSLAEE-----SPDLTQHPAVSVGGLEGESALTFLF--DLGA 241

QY 314 NCLGLDPVHQDDGPFSGHSGFLSAAALDADRYLLE-----SOLLASPNASTV 360


```

Db      242 S-IKHEPTHDLTAPLSDDDFRLLFNGDSLSLESDSSLEDDGFAPDVLDSGDLNFAFPDMSV 300
QY      361 DDYLAGDSAACTNPLPSDYDFDINDFLTDANHAAYDIVAASNYAAADRELDELHDP 420
Db      301 DFD-----TEPVTLLE-DLEQTNGLSDSASCKAASL----- 329
QY      421 ENQIPSRHSIQPOSGASSHGCDGGIAVG 450
Db      330 -----QPSHGASTSRCDDGGIAAG 348

RESULT 6
AAE15372
ID AAE15372 standard; protein; 350 AA.
XX
AC AAE15372;
XX
XX
DT 29-AUG-2003 (revised)
DT 07-MAR-2002 (first entry)
DE Aspergillus nidulans hacA protein.
XX
KW Heterologous protein secretion; unfolded protein response; UPR; lipase;
KW cellulase; carbohydrase; industry; purification; hacA protein.
XX
OS Emericella nidulans.
XX
FH Key Location/Qualifiers
FT Binding-site 53..116
FT /label= DNA-binding_domain
XX
XX US2001034045-A1.
XX
PD 25-OCT-2001.
XX
PF 23-MAR-2001; 2001US-00816277.
XX
PR 24-MAR-2000; 2000US-00534692.
XX
PA (GEMV ) GENENCOR INT INC.
XX
PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX WPI; 2002-033728/04.
DR N-PSDB; AAD24596.
XX
XX Increasing secretion of heterologous proteins e.g. lipase and cellulase
PT in eukaryotic cells useful in industry to increase production and
PT facilitate purification, by inducing an elevated unfolded protein
PT response.
XX
PS Claim 39; Fig 8; 56pp; English.
XX
XX The present invention relates to methods for increasing the secretion of
CC heterologous protein in eukaryotic cells by inducing an elevated unfolded
CC protein response (UPR). The method involves inducing the elevated UPR by
CC increasing the presence of proteins such as HAC1, HACA, PTC2 or IRE1 in
CC cells. The method and sequences are useful for increasing the secretion
CC of heterologous proteins (e.g. lipase, cellulase, carbohydrase) in
CC eukaryotic cells useful in industry to increase protein yields and to
CC facilitate purification. The present sequence is Aspergillus nidulans
CC hacA protein. (Updated on 29-AUG-2003 to standardise OS field)
XX
SQ Sequence 350 AA;
Query Match 19.1%; Score 437.5; DB 5; Length 350;
Best Local Similarity 32.4%; Pred.No.4e-27;
Matches 146; Conservative 51; Mismatches 122; Indels 131; Gaps 18;
QY 26 DNFTSL-----FADSTPTSLNPRDMTPDSVADIDSRLSVIPESQDAEDSHSATSATAP 80
Db 5 DRFSPVKMEDAFANSLPTTPSLVFLVTVSPADTSLRTKNV-----VAQTKP 51

```

```

QY      81 STSEKKPVKKRSGWOVLPEPKTNLPKRKRAKTEDEKQRRVERVLNRRAAQSRRERK 140
Db      52 --EEKKPAKKRSGWOGLPVPKTNLPKRKRAKTEDEKQRRVERVLNRRAAQSRRERK 109
QY      141 LEVEALEKRNKELETLINVOKTNLILVEELNFRFRSSGVVTRSSPLDSLQDSITLSQQ 200
Db      110 LEMEKLESEK-----IDMEQONQLLQRLQAEMENRLS-----QOVAQLSAE 153
QY      201 LFGSRDQGTMSNPEOSLMDQIMRSAAANTVTNPASLSPLPPIPSDKFQTKEDDEQADE- 259
Db      154 VRGSRH-----STPTSS-----SPASVSPTLTPTLFKQ-----EGDEVPLDRI 191
QY      260 --DEMEQOTWHTKEAAAAAKEKNKQSRVSTOSTORPAVSIGG-----DAAPVFESDAGA 313
Db      192 PPTPSVTVDYSPTLKPSLAE-----SPDLTOHPAVSVGLEGDESEALTLF--DLGA 241
QY      314 NCLGLDPVHQDDGPFPS---IGHFGLSAAALDADRYLLE-----SOLLASPNASTV 360
Db      242 S-IKHEPTHDLTAPLSDDDFRLLFNGDSLSLESDSSLEDDGFAPDVLDSGDLNFAFPDMSV 300
QY      361 DDYLAGDSAACTNPLPSDYDFDINDFLTDANHAAYDIVAASNYAAADRELDELHDP 420
Db      301 DFD-----TEPVTLLE-DLEQTNGLSDSASCKAASL----- 329
QY      421 ENQIPSRHSIQPOSGASSHGCDGGIAVG 450
Db      330 -----QPSHGASTSRCDDGGIAAG 348

RESULT 7
AAE15379
ID AAE15379 standard; protein; 386 AA.
XX
AC AAE15379;
XX
XX
DT 07-MAR-2002 (first entry)
DE Aspergillus niger var. awamori hacA protein #1.
XX
KW Heterologous protein secretion; unfolded protein response; UPR; lipase;
KW cellulase; carbohydrase; industry; purification; hacA protein.
XX
OS Aspergillus niger.
XX
XX US2001034045-A1.
XX
PD 25-OCT-2001.
XX
PF 23-MAR-2001; 2001US-00816277.
XX
PR 24-MAR-2000; 2000US-00534692.
XX
PA (GEMV ) GENENCOR INT INC.
XX
PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX WPI; 2002-033728/04.
DR N-PSDB; AAD24601.
XX
XX Increasing secretion of heterologous proteins e.g. lipase and cellulase
PT in eukaryotic cells useful in industry to increase production and
PT facilitate purification, by inducing an elevated unfolded protein
PT response.
XX
PS Claim 39; Fig 28; 56pp; English.
XX
XX The present invention relates to methods for increasing the secretion of
CC heterologous protein in eukaryotic cells by inducing an elevated unfolded
CC protein response (UPR). The method involves inducing the elevated UPR by
CC increasing the presence of proteins such as HAC1, HACA, PTC2 or IRE1 in
CC cells. The method and sequences are useful for increasing the secretion
CC of heterologous proteins (e.g. lipase, cellulase, carbohydrase) in
CC eukaryotic cells useful in industry to increase protein yields and to

```

CC facilitate purification. The present sequence is *Aspergillus niger* var.
CC awamori hacA protein
XX
SQ Sequence 386 AA;

[illegible]

RESULT 8	
ADSI2815	
ID	ADSI2815 standard; protein; 386 AA.
XX	AC
ADSI2815;	
XX	AC
XX	
DT	16-DEC-2004 (first entry)
XX	
DE	Aspergillus nidulans hcaA chaperone and foldase #3.
XX	
DE	unfolded protein response; UPR; HAC1; PTC2;
KW	unfolded protein response modulator; enzyme production; haca; chaperone;
KW	foldase; enzyme.
XX	
OS	Emericella nidulans.
XX	
US	US2004186070-A1.
PN	
XX	
PD	23-SEP-2004.
XX	
XX	
PF	15-SEP-2003; 2003US-00663450.
XX	
XX	24-MAR-2000; 2000US-00534692.
PR	
PR	23-MAR-2001; 2001US-00816277.
XX	
XX	
PA	(GEWV) GENENCOR INT INC.
XX	
PI	Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLÄ;
XX	
XX	WPI; 2004-707924/69.
DR	N-PSDB; ADSI2814.
DR	

Increasing the secretion of a heterologous protein, such as a thermostable PT or an industrial enzyme, in genetically modified eukaryotic cells by inducing an elevated unfolded protein response (UPR).

Example 12; SEQ ID NO 16; 83pp; English.

The invention describes a method of increasing the secretion of a heterologous protein in a eukaryotic cell, comprising inducing an elevated unfolded protein response (UPR). Also described are: an isolated nucleic acid encoding a HAC1 protein, where the HAC1 protein induces unfolded protein response and has less than 50% similarity to yeast HAC1 protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1 protein induces unfolded protein response and where the HAC1 protein comprises a DNA binding region that has greater than 70% similarity to the DNA binding region of filamentous fungi HAC1 protein; a protein having unfolded protein response inducing activity and having at least 70% similarity to a fully defined amino acid sequence of SEQ ID NO: 2, 4 or 16, or 12, 8, 14, 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12, 8, 14, respectively) as given in the specification; a protein having an amino acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated nucleic acid encoding a pTc2 protein that modulates unfolded protein response and has at least 70% similarity to a fully defined amino acid sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the specification; an isolated nucleic acid encoding a IRE1 protein that modulates unfolded protein response and has at least 60% similarity to a fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein having unfolded protein response modulating activity and having greater than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a heterologous nucleic acid encoding a protein having unfolded protein response modulating activity and a heterologous nucleic acid encoding a protein of interest to be secreted. The methods and compositions of genetically manipulating cells to have an elevated unfolded protein response (UPR) resulting in an increased capacity to produce secreted proteins, are useful in e.g. production of therapeutic antibodies, HA-tagged chaperone and foldase.

```

SQ Sequence 386 AA;
Query Match 18.9%; Score 434.5; DB 8; Length 186;
Best Local Similarity 32.3%; pred. NO. 8e 27;
Matches 147; Conservative 50; Mismatches 139; Indels 11; gaps 0;

QY 24 PGDNFTSLFADSTP-STLNPRDMTTPDSVADIDS ----- RLSSVIPESQDAFIDPES 72
DB 22 PGTTWPLGAPPASPVMETWPFVFLMWEAFSPVDSLAGSPTPELPLLTVPADTSLDSS-- 79
QY 73 HSTSATAPSTSEKPKPVKKRKSQGVLPBPKTNLPKRKAKTEDKEQRVERVLNRRAA 132
DB 80 -SVOAGETKABEKPKPVKKRKSQGVLPBPKTNLPKRKAKTEDKEQRVERVLNRRAA 138
QY 133 QSSRERKRLEVEALEKRKNEKETLIIINVOKTNLILVEELNRRPRSSGVVTRSSPTSDN 141
DB 139 OTSRRERKLENEKLENEK-----IOMEQNOFLIORSOMEAENNRIN 148
QY 193 DSITLSOOLFGRSGQGTWMSNPEQSLMDQIMKSAANPTVNPASLSPLPSIDSEKPTKEP 200
DB 183 QVAQLSAEVRGSR-GNT - PKPG-----SVVASPIILITLQFQ 206
QY 253 DE---EQADEDEEMQTHETKEAAAAAKSKSRVSTDSRTQRPASVIGG LAAVH 260
DB 221 DEIPLERIPPTPSIDTYSPTLRPSTLAE-----SSDVTQHFAVSVAQI RKE 243
QY 306 VFSDDAGANCLGLDPVHQDDGPFSGHSGFGLSAAALDAUKY---LLESQULASPNASIVU 313
DB 273 LF--DVGSN---PEPHAADD-----LAAPLSDDDFHRLFNVDSPVGSDDSSVLEDG 311
QY 363 ---DYLAGDSAACFTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAADRELDLEIHD 419
DB 318 FAFDVLDDGDDLSAF--PFDMSVDFD-----PESVCFEIE 325
QY 420 PENQIPSRHSIQ-----QPOSAGASSHCDDGGIAGV 450

```



```

Query Match      18.9%; Score 434; DB 8; Length 349;
Best Local Similarity 32.4%; Pred. No. 7.6e-27;
Matches 149; Conservative 50; Mismatches 109; Indels 152; Gaps 21;

QY 26 DNFTSL-----FADS---TFS-----TLNPRD---MMTPDSVADIDSRLSVIPESQDAEDD 70
DB 5 DRFPVKMEDAFANSPPTTSLVLPVLTVPADTSLRTKNVVAQ-----47
QY 71 ESHSTSATAPSTSEKPKVKKRKGQVLPKPTNLPKRAKTEDEKEQRRVERVLRNR 130
DB 48 -----TKP--BEKPKAKRKGQVLPKPTNLPKRAKTEDEKEQRRVERVLRNR 98
QY 131 AAQSSRRERKEVEALEKRNKELETLINVKTNLILVEELNFRFRSSGVTTRSSPLDS 190
DB 99 AAOQTSRRERKEVEALEKRNKELETLINVKTNLILVEELNFRFRSSGVTTRSSPLDS 143
QY 191 LQDSITLSQLFGSRDQGTMSNPESQSLMDQIMRASAANPTVNPASLSPSLPPIIDKKEFQTK 250
DB 144 -QOVAQLSAEVRGSRH-----STPSS-----SPASVSTLTPTLFKO-----180
QY 251 EDEEQADE---DEEMEQTWHETKEAAAKERNKSKOSRVSTSTQRPAYSIGG---DAA 303
DB 181 EGDEVPLDRIPPTPSTVDSPTLKPSLAEL-----SPDLTQHPAVSVGGLGDESA 232
QY 304 VPVFSDDAGANCLGLDPVHQDDGPFSS---IGHFGLSALDADRYLLE-----SQ 350
DB 233 LTLF--DLGAS--IKHEPTHTLDTAPLSDDDFRRLFNGLSDSSLESDDGFAFDVLDSDG 289
QY 351 LLAASPNASTVDDYLAGDSAACTNPLPSDYDPDINDFLTDDANHAAYDIVAASNYAAAD 410
DB 290 LSAFPDSDNVDF-----TEPVTLE-DLEQTNGLSDSNCKAASL-----328
QY 411 RELDLEIHPENQIPSRHSIQPQSGASHGCDGDIAGV 450
DB 329 -----QPSHGASRCDGQGIAG 347

RESULT 12
ID ABO43144 standard; protein; 409 AA.
XX ABO43144;
XX
XX
DT 23-SEP-2003 (first entry)
DE A. thaliana disease tolerance transcription factor, G1034.
XX
XX
KW Plant; transcription factor; disease resistance; transgenic;
KW plant breeding; pathogens resistance; pests; resistance.
XX
OS Arabidopsis thaliana.
XX
XX US2003046723-A1.
XX
XX 06-MAR-2003.
XX
XX 22-MAR-2000; 2000US-00533029.
XX
XX 22-MAR-2000; 2000US-00533029.
XX
XX (HEAR/) HEARD J.
XX (BROU/) BROUN P.
XX (RIEC/) RIECHMANN J L.
XX (KEDD/) KEDDIE J.
XX (PINE/) PINEDA O.
XX (ADAM/) ADAM L.
XX (SANA/) SANAH A.
XX (ZHAN/) ZHANG J.
XX (YUGG/) YU G.
XX (RATC/) RATCLIFFE O.
XX (PILG/) PILGRIM M.
XX (JIAN/) JIANG C.
XX (REUB/) REUBER L.

```

```

XX
PI PI Samaha R, Zhang J, Yu G, Ratcliffe O, Pilgrim M, Jiang C, Reuber L;
XX
XX WPI; 2003-521768/49.
DR N-PSDB; ACD98410.
XX
PT New transgenic plants comprising a recombinant gene that alters the
PT plant's disease tolerance or resistance, useful in plant breeding, e.g.
PT for generating plants with improved tolerance or resistance to diseases,
PT pests or pathogens.
XX
PS Claim 1; Page 100-101; 124pp; English.
XX
CC The invention relates to a transgenic plant, comprising a recombinant
CC polynucleotide that alters the plant's disease tolerance or resistance
CC when compared with the same trait of another plant lacking the
CC recombinant polynucleotide. The recombinant polynucleotide comprises a
CC nucleotide sequence, which encodes a polypeptide comprising at least 6
CC consecutive amino acids of any of 56 transcription factor proteins
CC appearing as ABO43093-ABO43148. Also included are altering the disease
CC tolerance or resistance of a plant (b); (a) transforming a plant with the
CC recombinant polynucleotide; (b) selecting the transformed plants; and (c)
CC identifying a transformed plant with an altered disease tolerance or
CC resistance), altering the expression levels of at least one gene in a
CC plant by transforming the plant with the recombinant polynucleotide,
CC altering a plant's trait (comprising: (a) providing a database sequence;
CC (b) comparing the database sequence with the polypeptide or
CC polynucleotide cited above; (c) selecting a database sequence that meets
CC the selected sequence criteria; and (d) transforming the selected
CC database sequence in the plant) and altering a plant's trait (comprising:
CC (a) providing a test polynucleotide; (b) hybridizing the test
CC polynucleotide at low stringency with the recombinant polynucleotide
CC cited above; and (c) transforming the hybridizing test polynucleotide in
CC a plant to alter a trait of the plant. The transgenic plant is useful in
CC plant breeding, particularly for generating plants with improved
CC tolerance or resistance to diseases. The plants have commercial utility
CC for increasing tolerance or resistance to pathogens and pests. The
CC present sequence is an Arabidopsis thaliana transcription factor of the
CC invention
XX
XX Sequence 409 AA;
XX
Query Match      18.7%; Score 429.5; DB 7; Length 409;
Best Local Similarity 38.2%; Pred. No. 2.2e-26;
Matches 116; Conservative 39; Mismatches 70; Indels 79; Gaps 10;

QY 11 KFEASPA---ESPLSAPGDNFTSLFADSTSTLNPRDMWTPDSVADIDSRLSVIPESOD 66
DB 8 KYENSPAETWVESFVSTPSSFHNPLFD---NNLNPVDFGFS-----QS 48
QY 67 AEDDESHTSATAPSTSEKPKVKKRKGQVLPKPTNLPKRAKTEDEKEQRRVERVYL 126
DB 49 FDRDYNFNGSLGSLNLPKPKPKKKGQVLPKPTNLPKRAKTEDEKEQRRVERVYL 108
QY 127 RNRAAQSSRRERKEVEALEKRNKELETLINVKTNLILVEELNFRFRSSGVTTRSSPLDS 176
DB 109 RNRAAQSSRRERKEVEALEKRNKELETLINVKTNLILVEELNFRFRSSGVTTRSSPLDS 168
QY 177 SSGVTRSSPLDSLQDS---ITLSQQLFGSRDQGTMSNPESQSLMDQIMRASAANPTVNP 232
DB 169 TNFLSYSDSSTPDISEDLSQLSPLTSPKQLFNAQD---ELCRP---ISPSIGPLTGRVDP 223
QY 233 ASLSPSLPPIIDKKEFQTKKEEEDQADEEMEQTWHETKEAAAKERNKSKOSRVSTSTQ 292
DB 224 STLSP-----KSLSSPDSSNSN-----SSDMTQ 246
QY 293 RPAV 296
DB 247 HPAV 250
XX
XX RESULT 13

```


PR 17-DEC-2002; 2002US-0434166P.
 PR 25-FEB-2003; 2003US-00374780.
 XX (ZHAN/) ZHANG J.
 PA (FROM/) FROMM M E.
 PA (HEAR/) HEARD J E.
 PA (RIEC/) RIECHMANN J L.
 PA (ADAM/) ADAM L J.
 PA (BROU/) BROUN P E.
 PA (PINE/) PINEDA O.
 PA (REUB/) REUBER T L.
 PA (KEDD/) KEDDIE J S.
 PA (YUGG/) YU G.
 PA (JIANG/) JIANG C.
 PA (SAMA/) SAMAHA R S.
 PA (PILG/) PILGRIM M L.
 PA (CREE/) CREELMAN R A.
 PA (DUBE/) DUBELL A N.
 PA (RATC/) RATCLIFFE O.
 PA (KUMI/) KUMIMOTO R.
 PA (SHER/) SHERMAN B K.
 XX
 PI Zhang J, Fromm ME, Heard JE, Riechmann JL, Adam LJ, Broun PE;
 PI Pineda O, Reuber TL, Keddie JS, Yu G, Jiang C, Samaha RS;
 PI Pilgrim ML, Creelman RA, Dubell AN, Ratcliffe O, Kumimoto R;
 PI Sherman BK;
 XX
 DR WPI; 2004-225755/21.
 DR N-PSDB; ADO02270.
 XX
 PT New transgenic plant, useful in developing phenotypes with altered or
 PT improved characteristics or traits.
 XX
 PS Claim 1; SEQ ID NO 684; 213pp; English.
 XX
 CC The invention relates to a transgenic plant comprises a recombinant
 CC polynucleotide having a polynucleotide sequence or its complementary
 CC sequence comprising a sequence encoding a polypeptide, that initiates
 CC transcription (i.e. a transcription factor) from Arabidopsis, Soybean,
 CC Rice, Rape or Corn, comprising any of the sequences appearing as ADO01588
 CC ADO03527 or ADO03530-ADO03559. Also included are using a transgenic
 CC plant to grow a progeny plant, an expression cassette (comprising a
 CC constitutive, inducible or tissue-specific promoter and a recombinant
 CC polynucleotide described above), a host cell comprising the expression
 CC cassette, producing a modified plant having a modified trait, identifying
 CC a factor that is modulated by or interacts with a polypeptide encoded by
 CC the polynucleotide sequence and identifying at least one downstream
 CC polynucleotide sequence that is subject to a regulatory effect of any of
 CC the polypeptides encoded by the polynucleotide described above. The
 CC transgenic plant is useful for producing a plant that has an altered
 CC trait e.g. an enhanced tolerance to abiotic stress (increased tolerance
 CC to chilling, germination in cold conditions, freezing tolerance, tolerance
 CC to heat, tolerance to drought, tolerance to osmotic stress, tolerance to
 CC salt, tolerance to phosphate limitation, tolerance to potassium
 CC limitation, decreased sensitivity to nitrogen limitation), altered
 CC hormone sensitivity, reduced sensitivity to abscisic acid, an altered
 CC response to ethylene, disease resistance, altered susceptibility to
 CC Botrytis, altered susceptibility to Fusarium, altered susceptibility to
 CC Erysiphe, altered susceptibility to Pseudomonas syringae, altered
 CC susceptibility to Sclerotinia, altered sugar sensing, improved seed
 CC germination and seedling vigor, early flowering, late flowering, extended
 CC period of flowering, an inflorescence architectural change, a change in
 CC stem bifurcations, a lack of a shoot meristem, reduced meristem cell
 CC differentiation, altered phyllotaxy, altered branching pattern, reduced
 CC apical dominance, reduced trichome density, ectopic trichome development,
 CC altered trichome development, altered stem morphology, increased root
 CC growth, increased root hair, altered seed development, altered cell
 CC proliferation/cell differentiation, premature senescence, delayed
 CC senescence, lethality, increased necrosis, an increase in seedling or
 CC plant size, decreased plant size, a change in leaf morphology, increased
 CC altered leaf development, increased leaf size and mass, glossy leaves,
 CC leaf cell expansion, change in seed morphology, altered seed coloration,
 CC increased seed size, decreased seed size, altered seed shape, change in

CC leaf biochemistry, increased leaf wax, an alteration in leaf prenyl lipid
 CC content, increased leaf insoluble sugars, decreased leaf fatty acid
 CC sugars, increased leaf anthocyanins, an alteration of leaf fatty acid
 CC content, an alteration of leaf glucosinolate content, change in seed
 CC biochemistry, an increase in seed oil content, decrease in seed oil
 CC content, increase in seed fatty acid content, decrease in seed fatty acid
 CC content, increase in seed protein content, decrease in seed protein
 CC content, alteration in seed prenyl lipid content, increase in seed
 CC sterols, upregulation of genes involved in secondary metabolism, increase
 CC in root anthocyanins, increase in plant anthocyanins, and alteration in
 CC light response or shade avoidance. The present sequence represents a
 CC thalacress transcription factor of the invention.
 XX
 SQ Sequence 409 AA;

Query Match 18.7%; Score 429.5; DB 8; Length 409;
 Best Local Similarity 38.2%; Pred. No. 2.2e-26;
 Matches 116; Conservative 39; Mismatches 70; Indels 79; Gaps 10;
 QY 11 KFEASPA-----ESFLSAPGDNFTSLFADSTSTLNPDMTPDSVADISRLSVIPESQD 66
 DB 8 KYENSPAETWVESFVSTPSSFHNPFLFD---NNLNPDVGFSP-----OS 48
 QY 67 AEDDESHTSATPSTSEKKPVKKKSGQVLPEPKTNLPKRKRAKTEDEKEQRRVERVL 126
 DB 49 FDRDYNFGSUSGLNLPKKPKIKKSGWQQQLPEPKTNLPKRKRAKTEDEKEQRRVERVL 108
 QY 127 RNRAAQSSRRKRLEVAL-----EKRNKELETLINQKTNLILVEELNR---FRR 176
 DB 109 RNRAAQSSRRKRQVEALEVEKAIERKNWDLMEAKYLLQQLKLRASGYNK 168
 QY 177 SSGVVTRSSSLDLSQDS-----ITLSQQLFGSRDQGTMSNPQSLMDQIMRRAAAPTVP 232
 DB 169 TNFLSYSDSTPDISEDSQLSPLTFSKQLFNAQD--ELCRP---ISPSIGLTSRTVDP 223
 QY 233 ASLSPLSLPISDKERQTKKEEDEEQADEMEQTHWETKRAAAKESKQSRVSTDSQ 292
 DB 234 STLSP-----KSLSPPSSNSN-----SSDMTQ 246
 QY 293 RPAV 296
 DB 247 HPAV 250

RESULT 15
 AAB82977
 ID AAB82977 standard; protein; 342 AA.

XX AC AAB82977;
 XX DT 11-SEP-2003 (revised)
 XX DT 21-DEC-2001 (first entry)
 XX DE Aspergillus niger hacA, involved in unfolded protein response.
 XX KW hacA; transcription factor; unfolded protein response; protein secretion.
 XX OS Aspergillus awamori.
 XX FH Key Location/Qualifiers
 XX FT Domain 45..109
 XX FT /label= DNA binding domain
 XX PN WO200172783-A2.
 XX PD 04-OCT-2001.
 XX PF 23-MAR-2001; 2001WO-US009401.
 XX PR 24-MAR-2000; 2000US-00534692.
 XX PA (GENV) GENENCOR INT INC.
 XX

Search completed: November 23, 2005, 03:18:37
Job time : 179.139 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:12:51 ; Search time 6.65899 Seconds
(without alignments)
844.265 Million cell updates/sec

Title: US-10-663-450-60

Perfect score: 342

Sequence: 1 KSTLPKRAKTKKEQRR.....CSLEENLNSVNLEKLADHE 68

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/1/iaa/5 COMB.pdp.*
- 2: /cgn2_6/ptodata/1/iaa/6 COMB.pdp.*
- 3: /cgn2_6/ptodata/1/iaa/H COMB.pdp.*
- 4: /cgn2_6/ptodata/1/iaa/PTUS COMB.pdp.*
- 5: /cgn2_6/ptodata/1/iaa/RE COMB.pdp.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	214	62.6	248	2	US-09-248-796A-18840
2	178.5	52.2	409	2	US-09-533-029-104
3	115	33.6	23	2	US-09-831-642-72
4	110	32.2	143	2	US-09-840-211A-1930
5	88.5	25.9	327	1	US-07-684-965-6
6	88.5	25.9	327	2	US-09-215-098-5
7	88.5	25.9	327	2	US-09-252-658-6
8	88.5	25.9	327	2	US-09-949-016-11202
9	87.5	25.6	68	2	US-09-299-495F-7
10	87.5	25.6	69	1	US-08-630-011A-7
11	87.5	25.6	92	2	US-09-299-495F-9
12	87.5	25.6	93	1	US-08-630-011A-9
13	86.5	25.3	315	2	US-09-949-016-7832
14	86	25.1	102	2	US-09-640-211A-1046
15	85.5	25.0	501	2	US-09-949-016-11281
16	84.5	24.7	521	1	US-08-721-684C-2
17	84.5	24.7	521	1	US-09-005-970-2
18	84.5	24.7	521	2	US-09-407-715-2
19	84	24.6	208	2	US-09-949-016-9078
20	84	24.6	295	1	US-08-343-443B-107
21	83	24.3	371	2	US-09-148-545-259
22	83	24.3	371	2	US-09-621-011-259
23	82.5	24.1	360	1	US-08-319-866-2
24	82.5	24.1	360	2	US-08-809-917-2
25	82.5	24.1	360	2	US-09-419-371-2
26	80	23.4	214	2	US-09-640-211A-1088
27	79.5	23.2	54	1	US-08-319-866-4

28	79.5	23.2	54	2	US-08-809-917-4	Sequence 4, Appli
29	79.5	23.2	54	2	US-09-419-371-4	Sequence 4, Appli
30	78.5	23.0	54	1	US-08-319-866-5	Sequence 5, Appli
31	78.5	23.0	54	2	US-08-809-917-5	Sequence 5, Appli
32	78.5	23.0	54	2	US-09-419-371-5	Sequence 5, Appli
33	78.5	23.0	351	2	US-09-870-089B-2	Sequence 2, Appli
34	78.5	23.0	362	2	US-09-949-016-10923	Sequence 10923, A
35	78	22.8	61	2	US-09-416-050A-26	Sequence 26, Appl
36	78	22.8	61	2	US-09-664-800-26	Sequence 26, Appl
37	78	22.8	61	2	US-09-665-309-26	Sequence 26, Appl
38	78	22.8	61	2	US-09-661-569-26	Sequence 26, Appl
39	78	22.8	307	2	US-09-267-031-6	Sequence 6, Appli
40	77.5	22.7	54	1	US-08-319-866-6	Sequence 6, Appli
41	77.5	22.7	54	2	US-08-809-917-6	Sequence 6, Appli
42	77.5	22.7	54	2	US-09-419-371-6	Sequence 6, Appli
43	77.5	22.7	71	2	US-09-270-767-35316	Sequence 35316, A
44	77.5	22.7	71	2	US-09-270-767-50533	Sequence 50533, A
45	76.5	22.4	196	2	US-09-640-211A-881	Sequence 881, App

ALIGNMENTS

RESULT 1

US-09-248-796A-18840
; Sequence 18840, Application US/09248796A
; Patent No. 6747137

; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 18840

; LENGTH: 248

; TYPE: PRT

; ORGANISM: Candida albicans

; FEATURE:

; NAME/KEY: UNSURE

; LOCATION: (221)

; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown

US-09-248-796A-18840

Query Match 62.6%; Score 214; DB 2; Length 248;

Best Local Similarity 73.3%; Pred. No. 5e-16;

Matches 44; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY 1 KSTLPKRAKTKKEQRRERILNRRRAAHSREKRLHQLYLRKCSLEENLNSVN 60

DB 72 KSTLPKRAKTKKEQRRERILNRRRAAHSREKRLHQLYLRKCSLEENLNSVN 131

RESULT 2

US-09-533-029-104

; Sequence 104, Application US/095333029

; Patent No. 6664446

; GENERAL INFORMATION:

; APPLICANT: Heard, Jacqueline

; APPLICANT: Broun, Pierre

; APPLICANT: Riechmann, Jose-Luis

; APPLICANT: Keddite, James

; APPLICANT: Pineda, Omalra

; APPLICANT: Adam, Luc

; APPLICANT: Samaha, Raymond

; APPLICANT: Zhang, James

; APPLICANT: Yu, Guo-Liang

; APPLICANT: Ratcliffe, Oliver

US-09-252-658-6

Query Match 25.9%; Score 88.5; DB 2; Length 327;
Best Local Similarity 33.9%; Pred. No. 0.047;
Matches 20; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

Qy 6 PRKAKTKKEKQRIIRILNRRAAHQSKREKKRLHLOYLERKCSLLENLLNSVNLKL 64
Db 258 PALPTQPAEAPRKREVLKMKREARECRKKKYYKVCLENRVAVLEN-QNKTLLIBEL 315

RESULT 8

US-09-949-016-11202
; Sequence 11202, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11202
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11202

Query Match 25.9%; Score 88.5; DB 2; Length 327;
Best Local Similarity 33.9%; Pred. No. 0.047;
Matches 20; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

Qy 6 PRKAKTKKEKQRIIRILNRRAAHQSKREKKRLHLOYLERKCSLLENLLNSVNLKL 64
Db 258 PALPTQPAEAPRKREVLKMKREARECRKKKYYKVCLENRVAVLEN-QNKTLLIBEL 315

RESULT 9

US-09-299-495F-7
; Sequence 7, Application US/09299495F
; Patent No. 6361968
; GENERAL INFORMATION:
; APPLICANT: VINSON, Charles R.
; KRYLOV, Dmitry
; TITLE OF INVENTION: EXTENSION OF A PROTEIN-PROTEIN
; INTERACTION SURFACE TO INACTIVATE THE FUNCTION OF A
; CELLULAR PROTEIN
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS WORD 97
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/299,495F
; FILING DATE: 26-Apr-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/001,654

US-09-252-658-6

Query Match 25.9%; Score 88.5; DB 2; Length 327;
Best Local Similarity 33.9%; Pred. No. 0.047;
Matches 20; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

Qy 6 PRKAKTKKEKQRIIRILNRRAAHQSKREKKRLHLOYLERKCSLLENLLNSVNLKL 64
Db 258 PALPTQPAEAPRKREVLKMKREARECRKKKYYKVCLENRVAVLEN-QNKTLLIBEL 315

RESULT 7

US-09-252-658-6
; Sequence 6, Application US/09252658
; Patent No. 6251667
; GENERAL INFORMATION:
; APPLICANT: HABENER, JOEL F.
; APPLICANT: HOFFLER, JAMES P.
; TITLE OF INVENTION: A CAMP-RESPONSIVE TRANSCRIPTIONAL
; TITLE OF INVENTION: ENHANCER BINDING PROTEIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVE., NW, STE. 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/252,658
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/684,965
; FILING DATE: 22-MAY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US89/05234
; FILING DATE: 20-NOV-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/272,980
; FILING DATE: 18-NOV-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Cimbala, Michele A
; REGISTRATION NUMBER: 33,851
; REFERENCE/DOCKET NUMBER: 0609.1990002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2600
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```
/
/ FILING DATE: 31-JUL-1995
/ APPLICATION NUMBER: 60/018,496
/ FILING DATE: 29-MAY-1996
/ APPLICATION NUMBER: 08/690,011
/ FILING DATE: 31-JULY-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Serunian, Leslie A.
/ REGISTRATION NUMBER: 35,353
/ REFERENCE/DOCKET NUMBER: 2026-4199US2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212)758-4800
/ TELEFAX: (212)751-6849
/ INFORMATION FOR SEQ ID NO: 7:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 68 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: unknown
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-299-495F-7

Query Match 25.6%; Score 87.5; DB 2; Length 68;
Best Local Similarity 37.3%; Pred. No. 0.011;
Matches 19; Conservative 15; Mismatches 16; Indels 1; Gaps 1;

QY 14 EEKEQRRIERILNRRAAHQSKREKKRLHLYLERKCSLLENLLNSVNLK 64
Db ||::||::||::||::||::||::||::||::||::||::||::||::||:|:|
7 EEAAKREVLMLKMRRAARECRKKKYYVKCLENRVAVLEN-QNKTLIBEL 56

RESULT 10
US-08-690-011A-7
/ Sequence 7, Application US/08690011A
/ Patent No. 5942433
/ GENERAL INFORMATION:
/ APPLICANT: VINSON, Charles R.
/ TITLE OF INVENTION: EXTENSION OF A PROTEIN-PROTEIN
/ TITLE OF INVENTION: INTERACTION SURFACE TO INACTIVATE THE FUNCTION OF A
/ TITLE OF INVENTION: CELLULAR PROTEIN
/ NUMBER OF SEQUENCES: 60
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
/ STREET: 345 Park Avenue
/ CITY: New York
/ STATE: NY
/ COUNTRY: USA
/ ZIP: 10154-0053
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30B
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/690,011A
/ FILING DATE: 31-JUL-1996
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 60/001,654
/ FILING DATE: 31-JUL-1995
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 60/018,496
/ FILING DATE: 29-MAY-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Serunian, Leslie A.
/ REGISTRATION NUMBER: 35,353
/ REFERENCE/DOCKET NUMBER: 2026-4199US1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212)751-6849
/ TELEFAX: (212)751-6849
/ INFORMATION FOR SEQ ID NO: 7:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 69 amino acids
```

```
/
/ TYPE: amino acid
/ STRANDEDNESS: unknown
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
US-08-690-011A-7

Query Match 25.6%; Score 87.5; DB 1; Length 69;
Best Local Similarity 37.3%; Pred. No. 0.011;
Matches 19; Conservative 15; Mismatches 16; Indels 1; Gaps 1;

QY 14 EEKEQRRIERILNRRAAHQSKREKKRLHLYLERKCSLLENLLNSVNLK 64
Db ||::||::||::||::||::||::||::||::||::||::||::||::||:|:|
7 EEAAKREVLMLKMRRAARECRKKKYYVKCLENRVAVLEN-QNKTLIBEL 56

RESULT 11
US-09-299-495F-9
/ Sequence 9, Application US/09299495F
/ Patent No. 6361968
/ GENERAL INFORMATION:
/ APPLICANT: VINSON, Charles R.
/ TITLE OF INVENTION: EXTENSION OF A PROTEIN-PROTEIN
/ TITLE OF INVENTION: INTERACTION SURFACE TO INACTIVATE THE FUNCTION OF A
/ TITLE OF INVENTION: CELLULAR PROTEIN
/ NUMBER OF SEQUENCES: 64
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
/ STREET: 345 Park Avenue
/ CITY: New York
/ STATE: NY
/ COUNTRY: USA
/ ZIP: 10154-0053
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: MS WORD 97
/ SOFTWARE: ASCII
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/299,495F
/ FILING DATE: 26-Apr-1999
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 60/001,654
/ FILING DATE: 31-JUL-1995
/ APPLICATION NUMBER: 60/018,496
/ FILING DATE: 29-MAY-1996
/ APPLICATION NUMBER: 08/690,011
/ FILING DATE: 31-JULY-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Serunian, Leslie A.
/ REGISTRATION NUMBER: 35,353
/ REFERENCE/DOCKET NUMBER: 2026-4199US2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212)758-4800
/ TELEFAX: (212)751-6849
/ INFORMATION FOR SEQ ID NO: 9:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 92 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: unknown
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-299-495F-9

Query Match 25.6%; Score 87.5; DB 2; Length 92;
Best Local Similarity 37.3%; Pred. No. 0.015;
Matches 19; Conservative 15; Mismatches 16; Indels 1; Gaps 1;

QY 14 EEKEQRRIERILNRRAAHQSKREKKRLHLYLERKCSLLENLLNSVNLK 64
Db ||::||::||::||::||::||::||::||::||::||::||::||::||:|:|
31 EEAAKREVLMLKMRRAARECRKKKYYVKCLENRVAVLEN-QNKTLIBEL 80
```

```
RESULT 12
US-08-690-011A-9
; Sequence 9, Application US/08690011A
; Patent No. 5942433
; GENERAL INFORMATION:
; APPLICANT: VINSON, Charles R.
; APPLICANT: KRYLOV, Dmitry
; TITLE OF INVENTION: EXTENSION OF A PROTEIN-PROTEIN
; TITLE OF INVENTION: INTERACTION SURFACE TO INACTIVATE THE FUNCTION OF A
; TITLE OF INVENTION: CELLULAR PROTEIN
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/690,011A
; FILING DATE: 31-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/001,654
; FILING DATE: 31-JUL-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/018,496
; FILING DATE: 29-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Serunian, Leslie A.
; REGISTRATION NUMBER: 35,353
; REFERENCE/DOCKET NUMBER: 2026-4199US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)759-4800
; TELEFAX: (212)751-6849
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-690-011A-9

Query Match 25.6%; Score 87.5; DB 1; Length 93;
Best Local Similarity 37.3%; Pred. No. 0.015;
Matches 19; Conservative 15; Mismatches 16; Indels 1; Gaps 1;

QY 14 EEKEORRIERILNRRAAHQSRKKRLHLYLQYLRKCSLLENLLNSVNLKL 64
Db 31 EEAAKREVLNKNREAAECRRKKYVKLCLENRVAVLEN-QNKTILBEL 80

RESULT 13
US-09-949-016-7832
; Sequence 7832, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11281
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Human

Query Match 25.6%; Score 87.5; DB 1; Length 93;
Best Local Similarity 37.3%; Pred. No. 0.015;
Matches 19; Conservative 15; Mismatches 16; Indels 1; Gaps 1;

QY 14 EEKEORRIERILNRRAAHQSRKKRLHLYLQYLRKCSLLENLLNSVNLKL 64
Db 31 EEAAKREVLNKNREAAECRRKKYVKLCLENRVAVLEN-QNKTILBEL 80

RESULT 14
US-09-640-211A-1046
; Sequence 1046, Application US/09640211A
; Patent No. 6833446
; GENERAL INFORMATION:
; APPLICANT: Wood, Marion
; APPLICANT: Shenk, Michael A.
; APPLICANT: McGrath, Annette
; APPLICANT: Glenn, Matthew
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Modification of Gene Transcription
; FILE REFERENCE: 11000.1021C1U
; CURRENT APPLICATION NUMBER: US/09/640,211A
; CURRENT FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 2368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1046
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-640-211A-1046

Query Match 25.1%; Score 86; DB 2; Length 102;
Best Local Similarity 40.8%; Pred. No. 0.025;
Matches 20; Conservative 9; Mismatches 20; Indels 0; Gaps 0;

QY 6 PRKRAKTKKEEQRIERILNRRAAHQSRKKRLHLYLQYLRKCSLLEN 54
Db 39 PRKRPSDLSAEKKEARAHNRIRIAAQNRSRDKRKQQTSLQORVIDLEN 87

RESULT 15
US-09-949-016-11281
; Sequence 11281, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11281
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Human
```

```
Query Match 25.3%; Score 86.5; DB 2; Length 315;
Best Local Similarity 37.3%; Pred. No. 0.076;
Matches 19; Conservative 15; Mismatches 16; Indels 1; Gaps 1;

QY 14 EEKEORRIERILNRRAAHQSRKKRLHLYLQYLRKCSLLENLLNSVNLKL 64
Db 254 EEATRKREVLNKNREAAECRRKKYVKLCLENRVAVLEN-QNKTILBEL 303

RESULT 14
US-09-640-211A-1046
; Sequence 1046, Application US/09640211A
; Patent No. 6833446
; GENERAL INFORMATION:
; APPLICANT: Wood, Marion
; APPLICANT: Shenk, Michael A.
; APPLICANT: McGrath, Annette
; APPLICANT: Glenn, Matthew
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Modification of Gene Transcription
; FILE REFERENCE: 11000.1021C1U
; CURRENT APPLICATION NUMBER: US/09/640,211A
; CURRENT FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 2368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1046
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-640-211A-1046

Query Match 25.1%; Score 86; DB 2; Length 102;
Best Local Similarity 40.8%; Pred. No. 0.025;
Matches 20; Conservative 9; Mismatches 20; Indels 0; Gaps 0;

QY 6 PRKRAKTKKEEQRIERILNRRAAHQSRKKRLHLYLQYLRKCSLLEN 54
Db 39 PRKRPSDLSAEKKEARAHNRIRIAAQNRSRDKRKQQTSLQORVIDLEN 87

RESULT 15
US-09-949-016-11281
; Sequence 11281, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11281
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Human
```

US-09-949-016-11281

Query Match 25.0%; Score 85.5; DB 2; Length 501;
Best Local Similarity 35.5%; Pred. No. 0.16;
Matches 22; Conservative 16; Mismatches 23; Indels 1; Gaps 1;

Qy 1 KSTLPKRKRKTKEEKEQRIERILNRRAAHQSRKRLHLOYLERKCSLLENLLNSVN 60
Db 333 QSTSGRRFRANEDPDEKRR-KFLERNFAAASRCQKQKRWVQSLKKAEDLSSINGOLQ 391

Qy 61 LE 62

Db 392 SE 393

Search completed: November 23, 2005, 03:27:34
Job time : 7.65899 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:05:10 ; Search time 29.2212 Seconds
(without alignments)
1641.819 Million cell updates/sec

Title: US-10-663-450-60

Perfect score: 342

Sequence: 1 KSTLPRKRAKTKKEQRR.....CSLLENLNSVNLKXLDHDE 68

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	342	100.0	230	1 HAC1 YEAST	P41546 saccharomyc
2	260	76.0	228	2 Q75BQ5 ASHGO	Q75BQ5 ashbya gosb
3	257	75.1	273	2 Q6CKQ1 KLJULA	Q6ckq1 kluyveromyc
4	214	62.6	357	2 Q5AA52 CANAL	Q5aa52 candida alb
5	211.5	61.8	260	2 Q6BQC2 DEBHA	Q6bqc2 debaryomyce
6	201	58.8	299	2 Q6CEV1 YARLI	Q6cev1 varrowia li
7	196	57.3	451	2 Q8TFF3 TRIRE	Q8tf3 trichoderma
8	193	56.4	430	2 Q7SHF0 NEUCR	Q7shf0 neurospora
9	182	53.2	556	2 Q5IKW8 MAGGR	Q5ikw8 magnaporthe
10	176.5	51.6	342	2 Q6W8X2 ASPNG	Q6w8x2 aspergillus
11	176.5	51.6	347	2 Q5AQN3 EMENI	Q5agn3 aspergillus
12	176.5	51.6	350	2 Q8TFU8 EMENI	Q8tfu8 emericalia
13	176.5	51.6	433	2 Q4WEY8 ASPFU	Q4wey8 aspergillus
14	173	50.6	329	2 Q6FLY3 CANGA	Q6fly3 candida gaa
15	157	45.9	429	2 Q4HTT5 GIBZE	Q4htt5 gibberella
16	124.5	36.4	168	1 HY5 ARATH	Q24646 arabidopsis
17	122	35.7	336	2 Q5VR11 ORYSA	Q5vr11 oryza sativ
18	121.5	35.5	69	2 Q5BL60 BRACM	Q5bl60 brassica ca
19	121.5	35.5	109	2 Q84XX8 BRARP	Q84xx8 brassica ra
20	110.5	32.3	322	2 Q04Z33 VICFA	Q04z33 vicia faba
21	109.5	32.0	321	2 Q8GRY7 LOTJA	Q8gry7 lotus japon
22	108.5	31.7	176	2 Q69XK6 ORYSA	Q69xk6 oryza sativ
23	107	31.3	120	2 Q682B6 ARATH	Q682b6 arabidopsis
24	107	31.3	149	1 HVH ARATH	Q8w191 arabidopsis
25	106.5	31.1	158	1 HY5 LYCES	Q9sm50 lycopersico
26	106	31.0	188	2 Q6ZHT8 ORYSA	Q6zht8 oryza sativ
27	104.5	30.6	322	2 Q39896 SOYBN	Q39896 glycine max
28	104.5	30.6	326	2 Q39895 SOYBN	Q39895 glycine max
29	103	30.1	646	2 Q6AU90 ORYSA	Q6au90 oryza sativ
30	102.5	30.0	141	2 Q677A7 OASPA	Q677a7 hyacinthus
31	96	28.1	267	2 Q7FTU1 ANOGA	Q7ftj1 anopheles 9

32	95.5	27.9	786	2	Q86AS2 DICDI	Q86as2 dictyosteli
33	95.5	27.9	787	2	Q554P0 DICDI	Q554p0 dictyosteli
34	95	27.8	372	2	Q54R29 DICDI	Q54r29 dictyosteli
35	92.5	27.0	327	2	Q6NVP9 XENTR	Q6nvp9 xenopus tro
36	92.5	27.0	703	2	Q4WC74 ASPFU	Q4wc74 aspergillus
37	91.5	26.8	306	2	Q8ION6 CAEEL	Q8ion6 caenorhabdi
38	91.5	26.8	314	2	Q8I4D8 CAEEL	Q8i4d8 caenorhabdi
39	91.5	26.8	324	2	Q8AVY5 XENLA	Q8avy5 xenopus lae
40	91.5	26.8	325	2	Q9U2I1 CAEEL	Q9u2i1 caenorhabdi
41	91.5	26.8	327	2	Q9U2I0 CAEEL	Q9u2i0 caenorhabdi
42	91.5	26.8	331	2	Q8I4D7 CAEEL	Q8i4d7 caenorhabdi
43	91.5	26.8	333	2	Q61BR7 CABBR	Q61br7 caenorhabdi
44	90.5	26.5	568	2	Q8LIB3 ORYSA	Q8lib3 oryza sativ
45	90	26.3	627	2	Q5BD44 EMENI	Q5bd44 aspergillus

ALIGNMENTS

RESULT 1
HAC1_YEAST
ID HAC1_YEAST STANDARD; PRT; 230 AA.
AC P41546; P87040;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE HAC1 protein.
GN Name=HAC1; Synonyms=ERN4, IRE2; OrderedLocusNames=YPL031W;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=S288c;
RX MEDLINE=95116316; PubMed=7816617;
RA Nojima H., Leem S.-H., Araki H., Sakai A., Nakashima N., Kanaoka Y., Ono Y.;
RT "Haci: a novel yeast bZIP protein binding to the CRE motif is a multicopy suppressor for cdc10 mutant of Schizosaccharomyces pombe.";
RL Nucleic Acids Res. 22:5279-5288(1994).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=S288c / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M., Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K., Yamazaki M.-A., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from Saccharomyces cerevisiae.";
RL Nat. Genet. 10:261-268(1995).
RN [3]
RP SEQUENCE REVISION TO 183-230.
RA Murakami Y.;
RT Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AB320;
RX MEDLINE=97222447; PubMed=9077435;
RA Mori K., Kawahara T., Yoshida H., Yanagi H., Yura T.;
RT "Signalling from endoplasmic reticulum to nucleus: transcription factor with a basic-leucine zipper motif is required for the unfolded protein-response pathway.";
RL Genes Cells 1:803-817(1996).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=97086687; PubMed=8932376; DOI=10.1093/nar/24.21.4222;
RA Nikawa J.-I., Akiyoshi M., Hirata S., Fukuda T.;
RT "Saccharomyces cerevisiae IRE2/HAC1 is involved in IRE1-mediated KAR2 expression.";
RL Nucleic Acids Res. 24:4222-4226(1996).
RN [6]
RP CHARACTERIZATION, AND ALTERNATIVE SPLICING.

```

RX MEDLINE=97053779; PubMed=8898193; DOI=10.1016/S0092-8674(00)81360-4;
RA Cox J.S., Walter P.;
RT "A novel mechanism for regulating activity of a transcription factor
RT that controls the unfolded protein response.";
RL Cell 87:391-404(1996).
CC -!- FUNCTION: Seems to be involved in the unfolded protein response
CC (UPR) pathway. Binds to the UPR element (UPRE) in the promoter of
CC UPR-regulated genes such as KAR2, PD11, EUG1 and PKB2; activates
CC the transcription of these genes.
CC -!- SUBUNIT: Homodimer.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=The level of each isoform is regulated by a splicing
CC event that occurs when the UPR is induced by IRE1;
CC Name=U;
CC IsoId=P41546-1; Sequences=Displayed;
CC Name=i;
CC IsoId=P41546-2; Sequences=VSP_000586;
CC Notes=Active and stable isoform which induces UPR;
CC -!- SIMILARITY: Belongs to the bZIP family.
CC -!- SIMILARITY: Contains 1 bZIP domain.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL; D26506; BAA05513.1; -; Genomic DNA.
DR EMBL; D50617; BAA24425.1; -; Genomic DNA.
DR EMBL; D86413; BAA19565.1; -; Genomic DNA.
DR PIR; S78571; S78571.
DR GeneOnline; 140124; -.
DR TRANSFAC; T02039; -.
DR Ensemble; YFL031W; Saccharomyces cerevisiae.
DR SGD; S00001863; HAC1.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0003704; F:specific RNA polymerase II transcription fa. . .; TAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0006644; P:phospholipid metabolism; TAS.
DR GO; GO:0045893; P:positive regulation of transcription, DNA-d. . .; IDA.
DR GO; GO:0006357; P:regulation of transcription from RNA polyme. . .; TAS.
DR GO; GO:0006990; P:unfolded protein response, positive regulat. . .; TAS.
DR InterPro; IPR011700; bZIP_2
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07716; bZIP_2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS00217; bZIP; FALSE NEG.
DR Activator; P900036; bZIP_BASIC; 1.
KW Nuclear protein; Transcription; Complete proteome; DNA-binding;
KW Unfolded protein response.
FT DOMAIN 67 81
FT DNA_BIND 29 61
FT VARSPLIC 221 230
FT -----
FT Leucine-zipper.
FT Basic motif.
FT AVITWTRKIQ --> EAQSLGSLFELNDFFITS (in
FT isoform 1).
FT FTID=VSP_000586.
FT RLCRVCRCFRVGPGRDFMGAAECLRRKMYQSRRLPYITIN
FT NLFDAVASPLADPLCDIAGNSLPFDNSIDLNDWRNPVIT
FT MTRKIQ --> ATLSPKSMRDSASDQETSWELOMFKTENPVE
FT STTLPADNNLNFDAVASRWQTHSATI (in Ref. 2).
FT RLCRVCRCFRVGPGRDFMGAAECLRRKMYQSRRLPYIT
FT --> ATLSPKSMRDSASDQETSWELOMFKTENPVESTLPVAV
FT DN (in Ref. 4).
FT SEQUENCE 230 AA; 26582 MW; 43073BCCCC4709B CRC64;
FT -----
Query Match 100.0%; Score 342; DB 1; Length 230;
Best Local Similarity 100.0%; Pred. No. 2,3e-23;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
FT -----
QY 1 KSTLPPRKAKTKEEQRIIRLNRRAAHQSRKKRLHLQYLERKCSLLENLNSVN 60
|||||

```

```

Db 23 KSTLPPRKAKTKEEQRIIRLNRRAAHQSRKKRLHLQYLERKCSLLENLNSVN 82
QY 61 LEKLADHE 68
Db 83 LEKLADHE 90
RESULT 2
Q75BQ5 ASHGO
ID Q75BQ5_ASHGO PRELIMINARY; PRT; 228 AA.
AC Q75BQ5;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE ACR216Cp.
GN Name=ACR216C;
OS Ashbya gossypii (Yeast) (Eremothecium gossypii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Eremothecium.
OX NCBI_TaxID=33169;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 10895;
RX PubMed=15001715; DOI=10.1126/science.1095781;
RA Dietrich F.S.; Voegelé S.; Brachat S.; Lerch A.; Gates K.; Steiner S.;
RA Mohr C.; Poehlmann R.; Luedi P.; Choi S.; Wing R.A.; Flavier A.;
RA Gaffney T.D.; Philippsen P.;
RT "The Ashbya gossypii genome as a tool for mapping the ancient
RT Saccharomycetes cerevisiae genome.";
RL Science 304:304-307(2004).
DR EMBL; AE016816; AAS51442.1; -; Genomic_DNA.
DR AGD; ACR216C; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; bZIP_2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07716; bZIP_2; 1.
DR PROSITE; PS00036; bZIP_BASIC; 1.
KW Complete proteome.
SQ SEQUENCE 228 AA; 24914 MW; 605A60D78633F038 CRC64;
Query Match 76.0%; Score 260; DB 2; Length 228;
Best Local Similarity 81.2%; Pred. No. 6.7e-16;
Matches 52; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
QY 1 KSTLPPRKAKTKEEQRIIRLNRRAAHQSRKKRLHLQYLERKCSLLENLNSVN 60
Db 14 KSTLPPRKAKTKEEQRIIRLNRRAAHQSRKKRLHLQYLERKCSLLENLNSVN 73
QY 61 LEKL 64
Db 74 LGAL 77
RESULT 3
Q6CKQ1 KLULA
ID Q6CKQ1_KLULA PRELIMINARY; PRT; 273 AA.
AC Q6CKQ1;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Kluyveromyces lactis strain NRRL Y-1140 chromosome F of strain NRRL Y-
DE 1140 of Kluyveromyces lactis.
GN OrderedLocustNames=KLUA0F08976g;
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=CBS 2359 / IFO 1267 / NRRL Y-1140 / WM37;
RX PubMed=15229592; DOI=10.1038/nature02579;

```



```

RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthonard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
RA Boisrame A., Boyer E., Fairhead C., Confanioli F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,
RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.-L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44 (2004).
DR EMBL; CR382126; CAG98196.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011616; BZIP_1.
DR Pfam; PF001170; bZIP_1; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS00036; BZIP BASIC; 1.
KW Complete proteome; DNA-binding; Nuclear protein.
SQ SEQUENCE 273 AA; 30429 MW; 97DFD7B1F8C6F5 CRC64;

Query Match 75.1%; Score 257; DB 2; Length 273;
Best Local Similarity 76.1%; Pred. No. 1.5e-15;
Matches 51; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

QY 1 KSTLPPRKAKTKEEQRIERILNRRRAAHSREKRLHLQYLERKCSLLENLNSVN 60
DB 15 KPTLPPRKAKTKEEQRIERILNRRRAAHSREKRLHLQYLERKCSLLENLNSVN 74
QY 61 LEKLADH 67
DB 75 LDILSEN 81

RESULT 4
Q5AA52 CANAL
ID Q5AA52_CANAL PRELIMINARY; PRT; 357 AA.
AC Q5AA52;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DE Hypothetical protein HAC1.
GN Name=HAC1; ORFNames=CaO19.2432, CaO19.9968;
OS Candida albicans SC5314.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=237561;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC PubMed=15123810; DOI=10.1073/pnas.0401648101;
RX Jones T., Federapfel N.A., Chibana H., Duncan J., Kalman S.,
RA Magee B.B., Newport G., Thorstenson Y.R., Agabian N., Magee P.T.,
RA Davis R.W., Scherer S.;
RT "The diploid genome sequence of Candida albicans.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:7329-7334 (2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC Dungan J., Kuo A., Newport G., Ian C.-Y., Iijima C., Adegbola O.,
RA Roberts J., Persson K., Donnelly S., Favoreto S., Tzung K.-W.,
RA Jones T., Scherer S., Agabian N.;
RT "Annotation of the Genome of Candida albicans.";
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AACQ0100041; EAK99529.1; -; Genomic DNA.
DR EMBL; AACQ0100040; EAK99617.1; -; Genomic_DNA.

```

```

DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
KW DNA-binding; Hypothetical protein; Nuclear protein.
SQ SEQUENCE 357 AA; 40127 MW; 8E1C63A2B4ECDAE CRC64;

Query Match 62.6%; Score 214; DB 2; Length 357;
Best Local Similarity 73.3%; Pred. No. 1.6e-11;
Matches 44; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY 1 KSTLPPRKAKTKEEQRIERILNRRRAAHSREKRLHLQYLERKCSLLENLNSVN 60
DB 54 KSTLPPRKAKTKEEQRIERILNRRRAAHSREKRLHLQYLERKCSLLENLNSVN 113

RESULT 5
Q6BQC2 DEBHA
ID Q6BQC2_DEBHA PRELIMINARY; PRT; 260 AA.
AC Q6BQC2;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Similarities with sp|P41546 Saccharomyces cerevisiae HAC1 protein.
DE OrderedLocusNames=DEHA0E071139g;
GN Debaryomyces hansenii (Yeast) (Torulaspora hansenii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.
OX NCBI_TaxID=4959;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 36239 / CBS 767;
RX PubMed=15295992; DOI=10.1038/nature02579;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthonard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
RA Boisrame A., Boyer E., Fairhead C., Confanioli F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.-P., Straub M.-L., Suleau A.,
RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.-L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44 (2004).
DR EMBL; CR382137; CAG87828.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; bZIP_2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07716; bZIP_2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS00217; BZIP; 1.
DR PROSITE; PS00036; BZIP BASIC; UNKNOWN 1.
KW Complete proteome; DNA-binding; Nuclear protein.
SQ SEQUENCE 260 AA; 29442 MW; 49B1819E393BFF6 CRC64;

Query Match 61.8%; Score 211.5; DB 2; Length 260;
Best Local Similarity 68.8%; Pred. No. 2e-11;
Matches 44; Conservative 7; Mismatches 8; Indels 5; Gaps 1;

QY 1 KSTLPPRKAKTKEEQRIERILNRRRAAHSREKRLHLQYLERKCSLLENLNSVN 60
DB 14 KSSLPPRKAKTKEEQRIERILNRRRAAHSREKRLHLQYLERKCSLLENLNSVN 68
QY 61 LEKL 64
DB 69 MSRL 72

```

Query Match 56.4%; Score 193; DB 2; Length 430;

```
Best Local Similarity 56.2%; Pred. No. 1.5e-09;
Matches 41; Conservative 12; Mismatches 14; Indels 6; Gaps 1;

Qy 1 KSTLPPPKRAKTKKEQRIERILNRRAAHQSRKKRLHLQYLERKCSLLENLNSV 56
    |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 105 KTNLPKRAKTEDEKEQRIERILNRRAAHQSRKKRLHLQYLERKCSLLENLNSV 164
    |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 57 --NSVLEKLADH 67
    |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 165 QINQOTLLQALREN 177
    |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 9
Q51KW8_MAGGR PRELIMINARY; PRT; 556 AA.
AC Q51KW8;
DT 13-SEP-2005 (TRENBLrel. 31, Created)
DT 13-SEP-2005 (TRENBLrel. 31, Last sequence update)
DE Hypothetical protein.
GN ORFNames-MG09010.4;
OS Magnaporthe grisea 70-15.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.
OX NCBI_TaxID=242507;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Birren B., Nusbaum C., Abebe A., Abouelleil A., Adekoya E.,
RA Ait-zahra M., Allen T., An P., Anderson M., Anderson S.,
RA Arachi H., Ambruster J., Bachanteang P., Baldwin J., Barry A.,
RA Bayul T., Blitshsteyn B., Bloom T., Bye J., Boguslavskiy L.,
RA Borowsky M., Boukhalter B., Brunache A., Butler J., Calixte N.,
RA Calvo S., Camarata J., Campo K., Chang J., Cheshatsang Y., Citroen M.,
RA Collymore A., Considine T., Cook A., Cooke P., Corum B., Cuomo C.,
RA David R., Dawoe T., Degray S., Dodge S., Dooley K., Dorje P.,
RA Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T., Engels R.,
RA Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,
RA Fitzgerald M., Foley K., Gage D., Galagan J., Gearin G., Gnerre S.,
RA Gnirke A., Goyette A., Graham J., Grandbois E., Gyaltsen K., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,
RA Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,
RA Jaffe D., Jones C., Kamal M., Kamat A., Kamvysselis M., Karlsson E.,
RA Kells C., Kieu A., Kisner P., Kodira C., Kulbokas E., Labutti K.,
RA Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,
RA Lindblad-toh K., Liu X., Lokyitsang T., Lokyitsang Y., Lucien O.,
RA Lui A., Ma L.J., Mabbitt R., Macdonald J., Maclean C., Major J.,
RA Manning J., Marabella R., Maru K., Matthews C., Mauceli E.,
RA McCarthy M., McDonough S., McGhee T., Meldrin J., Meneus L.,
RA Mesirov J., Mihalov A., Mihova T., Mikkelsen T., Mienga V., Moru K.,
RA Mozes J., Mulrain L., Munson G., Naylor J., News C., Nguyen C.,
RA Nguyen N., Nguyen T., Nicol R., Nielsen C., Nizzari M., Norbu C.,
RA Norbu N., O'donnell P., Okoawo O., O'leary S., Omotosho B.,
RA O'Neill K., Oman S., Parker S., Perrin D., Phunkhang P., Pignani B.,
RA Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond P.,
RA Retta R., Richardson S., Rhee C., Rodriguez J., Rogers J., Rogov P.,
RA Rutman M., Schupbach R., Seaman C., Settipalli S., Sharpe T.,
RA Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,
RA Spencer B., Stalker J., Stange-thomann N., Stavropoulos S.,
RA Steaton K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,
RA Tensing P., Tesfaye S., Theodore J., Thoultsang Y., Topham K.,
RA Tenkataran V., Tsonm N., Vallee D., Vassiliev H.,
RA Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,
RA Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembeck L.,
RA Zimmer A., Zody M., Lander E.;
RT "The genome sequence of Magnaporthe grisea.";
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Dean R., Mitchell T., Brown D., Pan H., Thon M.;
RA Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.

[3]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Zhu H., Blackmon B.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AACU01001606; EAA47880.1; -; Genomic_DNA.
DR InterPro; IPR011700; bZIP 2.
DR InterPro; IPR004827; TF bZIP.
DR Pfam; PF0716; bZIP 2; I.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS0217; bZIP; 1.
DR DNA-binding; Hypothetical protein; Nuclear protein.
KW SEQUENCE 556 AA; 60798 MW; CFDC4C8D2DC4963C CRC64;
SQ
    Query Match 53.2%; Score 182; DB 2; Length 556;
    Best Local Similarity 57.8%; Pred. No. 2e-08;
    Matches 37; Conservative 14; Mismatches 11; Indels 2; Gaps 1;

Qy 1 KSTLPPPKRAKTKKEQRIERILNRRAAHQSRKKRLHLQYLERKCSLLENLNSV 60
    |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 105 KTNLPKRAKTEDEKEQRIERILNRRAAHQSRKKRLHLQYLERKCSLLENLNSV 162
    |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 61 LEKL 64
    |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 163 FQKL 166
    |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 10
Q6W8X2_ASPNG PRELIMINARY; PRT; 342 AA.
AC Q6W8X2;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Transcription factor HACA.
GN Name=haca;
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Mulder H.J.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY303684; AAQ73495.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; bZIP 2.
DR InterPro; IPR004827; TF bZIP.
DR Pfam; PF0716; bZIP 2; I.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS0217; bZIP; 1.
DR PROSITE; PS00036; bZIP_BASIC; UNKNOWN 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 342 AA; 37148 MW; 623B3941A55C9C05 CRC64;
    Query Match 51.6%; Score 176.5; DB 2; Length 342;
    Best Local Similarity 55.9%; Pred. No. 3.9e-08;
    Matches 38; Conservative 12; Mismatches 17; Indels 1; Gaps 1;

Qy 1 KSTLPPPKRAKTKKEQRIERILNRRAAHQSRKKRLHLQYLERKCSLLENLNSV 60
    |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 64 KTNLPKRAKTEDEKEQRIERILNRRAAHQSRKKRLHLQYLERKCSLLENLNSV 122
    |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 61 LEKLADHE 68
    |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 123 LQRLSQME 130
    |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Sat Nov 26 13:50:41 2005

```
RESULT 11
Q5AQN3 EMENI PRELIMINARY; PRT; 347 AA.
AC Q5AQN3;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Hypothetical protein.
GN ORFNames=AN9397-2;
OS Aspergillus nidulans FGSC A4.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eutiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=227321;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FGSC A4;
RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavsky L.,
RA Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
RA Choepel Y., Collymore A., Cook A., Cooke P., Corum B., DeArellano K.,
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
RA Erickson J., Faro S., Ferreira P., FitzGerald M., Gage D., Galagan J.,
RA Gardyna S., Gnerre S., Graham L., Grand-Pierre N., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,
RA Jaffe D., Johnson R., Jones C., Kamat M., Kamat A., Karatas A.,
RA Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
RA Ma L.-J., Mabbitt R., MacLean C., Macdonald P., Major J., Manning J.,
RA Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,
RA Mihova T., Mlenga V., Murphy T., Naylor J., Nguyen C., Nicol R.,
RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neil D.,
RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,
RA Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,
RA Roman J., Schauer S., Schuback R., Seaman S., Severy P., Smirnov S.,
RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,
RA Talamas J., Tesfaye S., Theodore J., Topham K., Travers M.,
RA Vassiliev H., Venkataranan V.S., Viel R., Vo A., Wang S., Wilson B.,
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
RA Lander E.;
RT "Genomic Sequence of Aspergillus nidulans.";
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AACD01000172; EAA6464.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
KW DNA-binding; Hypothetical protein; Nuclear protein.
SQ SEQUENCE 347 AA; 37841 MW; A7CAAFEE761B3B1 CRC64;

Query Match 51.6%; Score 176.5; DB 2; Length 347;
Best Local Similarity 57.4%; Pred. No. 4e-08;
Matches 39; Conservative 11; Mismatches 17; Indels 1; Gaps 1;

QY 1 KSTLPKRAKTEKEQRRIILNRRAAHQSKREKHLHQLYERKCSLLENLSVN 60
DB 71 KTNLPKRAKTEKEQRRIILNRRAAHQSKREKHLHQLYERKCSLLENLSVN 129

QY 61 LEKLADHE 68
DB 130 LQRLAQME 137

RESULT 12
Q8TFU8 EMENI PRELIMINARY; PRT; 350 AA.
AC Q8TFU8;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Transcription factor.
GN Name=haa3;
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eutiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;

Query Match 51.6%; Score 176.5; DB 2; Length 347;
Best Local Similarity 57.4%; Pred. No. 4e-08;
Matches 39; Conservative 11; Mismatches 17; Indels 1; Gaps 1;

QY 1 KSTLPKRAKTEKEQRRIILNRRAAHQSKREKHLHQLYERKCSLLENLSVN 60
DB 71 KTNLPKRAKTEKEQRRIILNRRAAHQSKREKHLHQLYERKCSLLENLSVN 129

QY 61 LEKLADHE 68
DB 130 LQRLAQME 137

RESULT 13
Q4WEY8 ASPFU PRELIMINARY; PRT; 433 AA.
AC Q4WEY8;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE BZIP transcription factor (HacA), putative.
GN ORFNames=Afu3g04070;
OS Aspergillus fumigatus Af293.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eutiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=330879;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Af293;
RA Nierman W., Bain A., Anderson M.J., Wortman J., Kim H., Stanley J.,
RA Arroya J., Berriman M., Abe K., Archer D.B., Bermejo C., Bennett J.,
RA Bowyer P., Chen D., Collins M., Coulson N., Davies R., Dyer P.S.,
RA Farman M., Fedorova N., Fedorova N., Feldblyum T.V., Fischer R.,
RA Foster N., Fraser A., Garcia J.L., Garcia M.J., Goble A.,
RA Goldman G.H., Goni K., Griffith-Jones S., Gwilliam R., Haas B.,
RA Haas H., Harris D., Horiuchi H., Huang J., Humphrey S., Jimenez J.,
RA Keller N., Khouri H., Kitamoto K., Kobayashi T., Kulkarni R.,
RA Kumagai T., Lafton A., Latge J.-P., Li W., Lord A., Lu C.,
RA Majors W.H., May G.S., Miller B.L., Mohamoud Y., Molina M., Monod M.,
RA Mouyna I., Mulligan S., Murphy L., O'Neil S., Paulsen I.,
RA Penvalva M.A., Petrea M., Price C., Pritchard B.L., Quail M.A.,
RA Rabinowitz E., Rawlins N., Rajandream M.-A., Reichard U.,
RA Renauld H., Robson G.D., Rodriguez de Cordoba S., Rodriguez-Pena J.M.,
RA Ronning C.M., Rutter S., Salzberg S.L., Sanchez M.,
RA Sanchez-Ferrero J.C., Saunders D., Seeger K., Squares R., Squares S.,
RA Takeuchi M., Tekala P., Turner G., Vazquez de Aldana C.R., Weidman J.,
RA White O., Woodard J., Yu J.-H., Fraser C., Galagan J.E., Asai K.,
RA Machida M., Hall N., Barrell B., Denning D.W.;
RT "Genomic sequence of the pathogenic and allergenic filamentous fungus Aspergillus fumigatus.";
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AAHF01000010; EAL86689.1; -; Genomic_DNA.
```

```

DR InterPro: IPR011700; bZIP 2.
DR InterPro: IPR004827; TF_bZIP.
DR Pfam: PF07716; bZIP 2; 1.
DR SMART: SM00338; BRLZ; 1.
DR PROSITE: PS0217; bZIP; 1.
DR PROSITE: PS00036; bZIP_BASIC; UNKNOWN_1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 433 AA; 48908 MW; 3742DD0B43E05C74 CRC64;

Query Match      51.6%; Score 176.5; DB 2; Length 433;
Best Local Similarity 55.9%; Pred. No. 4.9e-08;
Matches 38; Conservative 12; Mismatches 17; Indels 1; Gaps 1;

QY 1 KSTLPKRAKTKSEKQRIERILNRRRAHQSRKRLHLYLQYLERKCSLLENLNSV 60
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
64 KTNLPKRAKTKSEKQRIERILNRRRAHQSRKRLHLYLQYLERKCSLLENLNSV 122
QY 61 LEKLADHE 68
DB :|:|:|:|
123 LQRLSQME 130

RESULT 14
Q6FLY3_CANGA PRELIMINARY; PRT; 329 AA.
AC Q6FLY3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Candida glabrata strain CBS138 chromosome K complete sequence.
GN OrderedLocusNamesCAGL0K125409;
OS Candida glabrata (yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
RN [1]_
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=15229592; DOI=10.1038/nature02579;
RA Lafontaine I., de Montigny J., Marc C., Neuveglise C., Talla E.,
RA Goffard N., Frangoul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
RA Boisrame A., Boyer J., Cattolico L., Confantoli F., de daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.-P., Straub M.-L., Suleau A.,
RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zenlou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.-L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
DR EMBL; CR380957; CAG61724.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR011616; bZIP 1.
DR InterPro: IPR004827; TF_bZIP.
DR Pfam: PF00170; bZIP; 1.
DR SMART: SM00338; BRLZ; 1.
DR PROSITE: PS0217; bZIP; 1.
DR PROSITE: PS00036; bZIP_BASIC; 1.
KW Complete proteome; DNA-binding; Nuclear protein.
SQ SEQUENCE 329 AA; 37162 MW; 469614BF37AF2282 CRC64;

Query Match      50.6%; Score 173; DB 2; Length 329;
Best Local Similarity 50.8%; Pred. No. 7.9e-08;
Matches 33; Conservative 19; Mismatches 9; Indels 4; Gaps 1;

QY 4 LPKRAKTKSEKQRIERILNRRRAHQSRKRLHLYLQYLERKCSLLENLNSV 59

```

```

Db 29 MPKRAKTKSEKQRIERILNRRRAHQSRKRLHLYLQYLERKCSLLENLNSV 88
QY 60 NLEK 64
Db 89 DIKSM 93

RESULT 15
Q4HTT5_GIBZE PRELIMINARY; PRT; 429 AA.
AC Q4HTT5;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=FG11623.1;
OS Gibberella zeae PH-1.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
OX NCBI_TaxID=229533;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PH-1;
RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavsky L.,
RA Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
RA Choepel Y., Collymore A., Cook A., Cooke P., Corum B., DeArellano K.,
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
RA Erickson J., Fero S., Ferreira P., FitzGerald M., Gage D., Galagan J.,
RA Gardyna S., Gnerre S., Graham L., Grand-Pierre N., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,
RA Jaffe D., Johnson R., Jones C., Kamal M., Kamat A., Karatas A.,
RA Kellis C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
RA Ma L.-J., Mabbitt R., Maclean C., Macdonald P., Major J., Manning J.,
RA Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,
RA Minova T., Mlenga V., Murphy T., Naylor J., Nguyen C., Nicol R.,
RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neil D.,
RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,
RA Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,
RA Roman J., Schauer S., Schupbach R., Seaman S., Severy P., Smirnov S.,
RA Smith C., Spencer B., Strange-Thomann N., Stojanovic N., Stubbs M.,
RA Talamas J., Tesfaye S., Theodore J., Topham K., Travers M.,
RA Vasiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
RA Lander E.;
RT "Fusarium graminearum genome sequence.";
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
KW Hypothetical protein.
SQ SEQUENCE 429 AA; 46630 MW; 04CCDE40BD392901 CRC64;

Query Match      45.9%; Score 157; DB 2; Length 429;
Best Local Similarity 47.9%; Pred. No. 2.9e-06;
Matches 34; Conservative 10; Mismatches 9; Indels 18; Gaps 1;

QY 1 KSTLPP-----RKRATKEEKEQRIERILNRRRAHQSRKRLHLYLQYLERKCSLLENLNSV 42
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100 KTNLPKRAKTKSEKQRIERILNRRRAHQSRKRLHLYLQYLERKCSLLENLNSV 159
QY 43 QYLERKCSLLE 53
Db 160 EALEKNGEQL 170

Search completed: November 23, 2005, 03:25:00
Job time : 30.2212 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:07:01 ; Search time 4.7788 Seconds
(without alignments)
1369.117 Million cell updates/sec

Title: US-10-663-450-60
Perfect score: 342
Sequence: 1 KSTLPFRKAKTKSEKEQRR.....CSLLENLLNSVNLKLDHHE 68

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80.*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	342	100.0	238	2 S78571	transcription fact
2	124.5	36.4	168	2 T50922	bZIP protein HY5 l
3	110.5	32.3	332	2 T12093	TGACG-motif bindin
4	104.5	30.6	322	2 T08592	TGACG-motif bindin
5	104.5	30.6	326	2 T08591	TGACG-motif bindin
6	91.5	26.8	331	2 T26807	hypothetical prote
7	91.5	26.8	333	2 T26808	hypothetical prote
8	89.5	26.2	515	2 A42140	box B-binding fact
9	89.5	26.2	516	2 A44494	CAMP-responsive el
10	88.5	25.9	325	2 S23007	CAMP response elem
11	88.5	25.9	326	2 A40120	CAMP-responsive en
12	88.5	25.9	327	2 A35769	CAMP response elem
13	88.5	25.9	327	2 S20955	regulatory protein
14	88.5	25.9	327	2 S22298	CAMP response elem
15	88.5	25.9	328	2 A35663	CAMP response elem
16	88.5	25.9	341	2 S42699	hypothetical prote
17	88.5	25.9	341	2 B35769	CAMP response elem
18	88.5	25.9	341	2 S03343	CAMP response elem
19	88.5	25.9	412	2 D86203	hypothetical prote
20	87.5	25.6	120	2 JX0307	cyclic AMP-respons
21	87	25.4	486	2 JC4028	activating transcr
22	86.5	25.3	108	2 A49317	inducible CAMP ear
23	86.5	25.3	229	2 A37944	CAMP response elem
24	86.5	25.3	278	2 S36101	CAMP response elem
25	86.5	25.3	344	2 JC5601	CAMP response elem
26	85.5	25.0	378	2 T07154	bZIP DNA-binding p
27	85.5	25.0	505	1 S05380	transcription fact
28	85	24.9	176	2 B90087	hypothetical prote
29	85	24.9	269	2 A46490	TCR V beta CRE-mot

30	85	24.9	360	2 T03373	probable G-box bin
31	84.5	24.7	327	2 S22299	CAMP response elem
32	84.5	24.7	393	2 S16321	light-induced prot
33	84	24.6	271	2 S12560	transcription fact
34	84	24.6	295	4 S36174	RNA binding protei
35	84	24.6	313	2 A34785	DNA-binding protei
36	84	24.6	358	2 C42026	cyclic AMP respons
37	84	24.6	389	1 A39429	CAMP response elem
38	84	24.6	448	2 A42026	CAMP response elem
39	84	24.6	456	2 B42026	cyclic AMP respons
40	83.5	24.4	297	2 T51273	promoter-binding f
41	83.5	24.4	600	2 T00759	hypothetical prote
42	83	24.3	688	2 T32750	hypothetical prote
43	82.5	24.1	246	2 T12585	DC3 promoter-bindi
44	82	24.0	483	2 S12741	transcription fact
45	82	24.0	620	2 T49067	transcription fact

ALIGNMENTS

RESULT 1

S78571
transcription factor HAC1 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YFL031w
C;Species: Saccharomyces cerevisiae
C;Date: 13-Sep-1998 #sequence revision 13-Sep-1998 #text_change 09-Jul-2004
C;Accession: S78571; S56223; S53578
R;Murakami, Y.
submitted to the Protein Sequence Database, January 1998
A;Reference number: S78570
A;Accession: S78571
A;Molecule type: DNA
A;Residues: 1-238 <MUR>
A;Cross-references: UNIPROT:P41546; UNIPARC:UPI0000157F61; EMBL:D50617; MIPS:YFL031w
A;Note: this is a revision to the sequence from reference S56186
R;Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sasanuma, S.I.;
submitted to the EMBL Data Library, May 1995
A;Description: Analysis of the nucleotide sequence of chromosome VI from Saccharomyces cerevisiae
A;Reference number: S56186
A;Accession: S56223
A;Molecule type: DNA
A;Residues: 1-191, 'RMOTHSATI', <MUM>
A;Cross-references: UNIPARC:UPI0000179B60; EMBL:D50617; NID:g836685; PID:g836723; MIPS:YFL031w
A;Note: this sequence has been revised in reference S78570
A;Note: this was believed to be the complete sequence of YFL031w
R;Nojima, H.; Leem, S.H.; Araki, H.; Sakai, A.; Nakashima, N.; Kanaoka, Y.; Ono, Y.
Nucleic Acids Res. 22, 5279-5288, 1994
A;Title: Hac1: a novel yeast bZIP protein binding to the CRE motif is a multicopy suppressor of the
A;Reference number: S53578; MUID:95116316; PMID:7616617
A;Accession: S53578
A;Molecule type: DNA
A;Residues: 1-142, 'RLCRPRVCRFRVGRDPFGMAECLRRMYQSRRLYPYTI', 183-220, 'AVITMTRKLQ' <NOJ>
A;Cross-references: UNIPARC:UPI000012C042; GB:D26506; NID:g633122; PIDN:BAA05513.1; PID:5513.1
C;Genetics:
A;Gene: SGD:HAC1; IRE2; ERN4
A;Cross-references: SGD:S0001863; MIPS:YFL031w
A;Map position: 6L
A;Introns: 221/1
C;Keywords: DNA binding; nucleus; transcription factor

Query Match 100.0%; Score 342; DB 2; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.8e-24;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	KSTLPFRKAKTKSEKEQRRIRILNRRRAAQSREKRLHLQYLRKCSLLENLSVN	60
DB	23	KSTLPFRKAKTKSEKEQRRIRILNRRRAAQSREKRLHLQYLRKCSLLENLSVN	82
QY	61	LEKLADHE	68
DB	83	LEKLADHE	90

RESULT 2

T50922
C; Species: Arabidopsis thaliana [mouse-ear cress]
C; Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004
C; Accession: T50922
R; Oyama, T.; Shimura, Y.; Okada, K.
submitted to the EMBL Data Library, July 1997
A; Description: The Arabidopsis HY5 gene encodes a bZIP protein that regulates stimulus-induced transcriptional activation of auxin-responsive genes
A; Reference number: Z25271
A; Accession: T50922
A; Status: preliminary; translated from GB/EMBL/DDBB
A; Molecule type: mRNA
A; Residues: 1-168 <OVA>
A; Cross-references: UNIPROT:O24646; UNIPARC:UIP000012CFCA; EMBL:AB005295; PIDN:BAA21116
A; Experimental source: Landsberg erecta

Query Match 36.4%; Score 124.5; DB 2; Length 168;
Best Local Similarity 47.5%; Pred. No. 0.00012;
Matches 28; Conservative 13; Mismatches 17; Indels 1; Gaps 1;

Qy 7 RKRATKEEKERIERILNRRAAHOSREKRLHLQLYLRKCGLLENLNLSVNLKLA 65
 ||||| :||:||||| :||:||||| :||:||||| :||:|||||
Dd 78 RKGRTPAEKNKKLLNNRVSAQAQRERRKAYLSELENVKDLEN-KNSELERIS 135

RESULT 3

T12093
TGACG-motif binding protein - fava bean
C; Species: Vicia faba (fava bean)
C; Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 31-Dec-2004
C; Accession: T12093
R; Wohlfarth, T.
submitted to the EMBL Data Library, May 1996
A; Reference number: Z17415
A; Accession: T12093
A; Status: preliminary; translated from GB/EMBL/DDBB
A; Molecule type: mRNA
A; Residues: 1-322 <WOH>
A; Cross-references: UNIPROT:O04234; UNIPARC:UIP000000A297E; EMBL:X97904
C; Superfamily: TGACG-motif-binding transcription factor

Query Match 32.3%; Score 110.5; DB 2; Length 322;
Best Local Similarity 42.4%; Pred. No. 0.0041;
Matches 25; Conservative 16; Mismatches 17; Indels 1; Gaps 1;

Qy 7 RKRATKEEKERIERILNRRAAHOSREKRLHLQLYLRKCGLLENLNLSVNLKLA 65
 ||||| :||:||||| :||:||||| :||:||||| :||:|||||
Dd 234 KKGRSPADKSKLKELLNNRVSAQAQRERRKAYLSDLETVRVDLEK-KNSCLKELKS 291

RESULT 4

T08592
TGACG-motif-binding protein STP2 - soybean
C; Species: Glycine max (soybean)
C; Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 31-Dec-2004
C; Accession: T08592
R; Cheong, Y.H.; Yoo, C.M.; Park, J.M.; Ryu, G.R.; Goekjian, V.H.; Nagao, R.T.; Key, J.L.
submitted to the EMBL Data Library, September 1995
A; Description: STP1 is a novel TGACG-binding factor with a zinc-finger motif and a bZIP domain
A; Reference number: Z16445
A; Accession: T08592
A; Status: preliminary; translated from GB/EMBL/DDBB
A; Molecule type: mRNA
A; Residues: 1-322 <CHE>
A; Cross-references: UNIPROT:Q39896; UNIPARC:UIP000000A61C8; EMBL:L28004; NID:S986966; PTD:
C; Genetics:
A; Gene: STP2
C; Superfamily: TGACG-motif-binding transcription factor

Query Match 30.6%; Score 104.5; DB 2; Length 322;

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-333 <WIL>

A;Cross-references: UNIPROT:Q9U210; UNIPARC:UPI000017A523; EMBL:AL032627; PIDN:CAB54382.

A;Experimental source: clone Y41C4A

C;Genetics:

A;Gene: CBSP:Y41C4A.4b

A;Introns: 24/3; 50/2; 81/3; 161/1; 230/1; 294/3

Query Match 26.8%; Score 91.5; DB 2; Length 333;
Best Local Similarity 38.5%; Pred. No. 0.24;
Matches 20; Conservative 15; Mismatches 16; Indels 1; Gaps 1;
QY 13 KEKEQRRIERILNRRAAHQSRKKRLHLQYLERKCSLLENLNSVNLEKL 64
DB 268 EDESRRKQVRLDKNREAKKCRKKKYYKCLNRRVSVLEN-QNKALIEEL 318

RESULT 8

A42140

box B-binding factor-2 - fruit fly (*Drosophila melanogaster*)

C;Species: *Drosophila melanogaster*

C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004

C;Accession: A42140; S24542

R;Abel, T.; Bhatt, R.; Maniatis, T.

Genes Dev. 6, 468-480, 1992

A;Title: A *Drosophila* CREB/ATF transcriptional activator binds to both fat body- and liv

A;Reference number: A42140; MUID:92192458; PMID:1532159

A;Accession: A42140

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-515 <ABE>

A;Cross-references: UNIPROT:P29747; UNIPARC:UPI00001283CE; EMBL:X64429; NID:g11063; PIDN

A;Note: sequence extracted from NCBI backbone (NCBIP:88160)

C;Genetics:

A;Gene: FlyBase:Bbbf2

A;Cross-references: FlyBase:FBgn0004848

C;Superfamily: CAMP-responsive element-binding transcription activator; fos/jun DNA-bind

C;Keywords: DNA binding; nucleus; transcription regulation

F;432-475/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 26.2%; Score 89.5; DB 2; Length 515;
Best Local Similarity 37.5%; Pred. No. 0.53;
Matches 24; Conservative 14; Mismatches 23; Indels 3; Gaps 2;

QY 6 PRKRATK-EEKEQRRIERILNRRAAHQSRKKRLHLQYLERKCSLLENLNSVNLE 62

DB 428 PQKLPTKAEKSLKIRRKIKNKISQAQSRKKKQYMDQLERRVILVTENHDYKKRLE 487

QY 63 KLAD 66

DB 488 GLEE 491

RESULT 9

A44494

CAMP-responsive element-binding transcription activator CREB-A - fruit fly (*Drosophila m*

C;Species: *Drosophila melanogaster*

C;Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004

C;Accession: A44494

R;Smolik, S.M.; Rose, R.E.; Goodman, R.H.

Mol. Cell. Biol. 12, 4123-4131, 1992

A;Title: A cyclic AMP-responsive element-binding transcriptional activator in *Drosophila*

A;Reference number: A44494; MUID:92375081; PMID:1508208

A;Accession: A44494

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-516 <SMO>

A;Cross-references: UNIPROT:P29747; UNIPARC:UPI0000078073; GB:M87038; NID:g157118; PID:g

A;Note: sequence extracted from NCBI backbone (NCBIN:111718, NCBIP:111719)

C;Genetics:

A;Gene: FlyBase:Creba

A;Cross-references: FlyBase:FBgn0004396

C;Superfamily: CAMP-responsive element-binding transcription activator; fos/jun DNA-bind
F;433-476/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 26.2%; Score 89.5; DB 2; Length 516;
Best Local Similarity 37.5%; Pred. No. 0.53;
Matches 24; Conservative 14; Mismatches 23; Indels 3; Gaps 2;

QY 6 PRKRATK-EEKEQRRIERILNRRAAHQSRKKRLHLQYLERKCSLLENLNSVNLE 62
DB 429 PQKLPTKAEKSLKIRRKIKNKISQAQSRKKKQYMDQLERRVILVTENHDYKKRLE 488

QY 63 KLAD 66

DB 489 GLEE 492

RESULT 10

S23007

CAMP response element-binding protein CREB.2 - bovine

C;Species: *Bos primigenius taurus* (cattle)

C;Date: 08-Jun-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004

C;Accession: S23007

R;Williams, L.; Kettmann, R.; Chen, G.; Portetelle, D.; Burny, A.; Derse, D.

DNA Seq. 1, 415-417, 1991

A;Title: Nucleotide sequence of the bovine cyclic-AMP responsive DNA binding protein (CRI

A;Reference number: S23007; MUID:92119333; PMID:1837490

A;Accession: S23007

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-325 <WIL>

A;Cross-references: UNIPROT:P27925; UNIPARC:UPI0000167BE3; EMBL:X57031; NID:g285; PIDN:C

C;Keywords: DNA binding; transcription regulation

F;259-302/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 25.9%; Score 88.5; DB 2; Length 325;
Best Local Similarity 33.9%; Pred. No. 0.44;
Matches 20; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

QY 6 PRKRATK-EEKEQRRIERILNRRAAHQSRKKRLHLQYLERKCSLLENLNSVNLEKL 64
DB 256 PALPTQAEAAKREVRMLKNREAAECRRKKKYYKCLNRRVAVLEN-QNKTLIEEL 313

RESULT 11

A40120

CAMP-responsive enhancer-binding protein CREB - human

C;Species: *Homo sapiens* (man)

C;Date: 27-Mar-1992 #sequence_revision 27-Mar-1992 #text_change 15-Mar-2004

C;Accession: A40120

R;Hoefler, J.P.; Meyer, T.E.; Yun, Y.; Jameson, J.L.; Habener, J.F.

Science 242, 1430-1433, 1988

A;Title: Cyclic AMP-responsive DNA-binding protein: structure based on a cloned placental

A;Reference number: A40120; MUID:89072714; PMID:2974179

A;Accession: A40120

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-326 <HOE>

A;Cross-references: UNIPARC:UPI000017A520; GB:M27691

F;260-303/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 25.9%; Score 88.5; DB 2; Length 326;
Best Local Similarity 33.9%; Pred. No. 0.44;
Matches 20; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

QY 6 PRKRATK-EEKEQRRIERILNRRAAHQSRKKRLHLQYLERKCSLLENLNSVNLEKL 64
DB 257 PALPTQAEAAKREVRMLKNREAAECRRKKKYYKCLNRRVAVLEN-QNKTLIEEL 314

RESULT 12

A35769

CAMP response element-binding protein A - human

C;Species: *Homo sapiens* (man)

C>Date: 05-Oct-1990 #sequence_revision 05-Oct-1990 #text_change 09-Jul-2004
A:Accession: A37340; A35769
R:Waechter, G.; Meyer, T.E.; Hoeffler, J.P.; Habener, J.F.
Trans. Assoc. Am. Physicians 103, 28-37, 1990
A:Title: Diversification of cyclic AMP-responsive enhancer binding proteins generated by
A:Reference number: A37340; PMID:92087371; PMID:1966745
A:Accession: A37340
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-327 <WAE>
A:Cross-references: UNIPROT:PI6220; UNIPARC:UPI000002A717; GB:S72459; NID:g240428; PIDN:
R:Berikowitz, L.A.; Gilman, M.Z.
Proc. Natl. Acad. Sci. U.S.A. 87, 5258-5262, 1990
A:Title: Two distinct forms of active transcription factor CREB (cAMP response element b
A:Reference number: A35769; MUID:90319091; PMID:2142528
A:Accession: A35769
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-327 <BER>
A:Cross-references: UNIPARC:UPI000002A717; GB:M34356; NID:g181042; PIDN:AAA35716.1; PID:
C:Keywords: alternative splicing; DNA binding; transcription regulation
F:261-304/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 25.9%; Score 88.5; DB 2; Length 327;
Best Local Similarity 33.9%; Pred. No. 0.44;
Matches 20; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

QY 6 PRKRAKTEEKQRRRIILNRRAAHQSRKKRLHLOYLERKCSLLENLLNSVNLKL 64
Db 258 PALPTQPAEEAARKREVRMLKMKREAAECRRKKKEYVKCLNRAVAVLEN-QNKTLIEEL 315

RESULT 13
S20955
regulatory protein CREB - mouse
N:Alternate names: cAMP responsive element-binding protein
C:Species: Mus musculus (house mouse)
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
C:Accession: S20955
R:Ruppert, S.; Cole, T.J.; Boshart, M.; Schmid, E.; Schuetz, G.
EMBO J. 11, 1503-1512, 1992
A:Title: Multiple mRNA isoforms of the transcription activator protein CREB: generation
A:Reference number: S20955; MUID:92224889; PMID:1532935
A:Accession: S20955
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-327 <RUP>
A:Cross-references: UNIPROT:Q01147; UNIPARC:UPI0000004066; EMBL:X67727
C:Genetics:
A:Introns: 38/3; 87/3; 121/2; 169/1; 230/1; 280/2
F:261-304/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 25.9%; Score 88.5; DB 2; Length 327;
Best Local Similarity 33.9%; Pred. No. 0.44;
Matches 20; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

QY 6 PRKRAKTEEKQRRRIILNRRAAHQSRKKRLHLOYLERKCSLLENLLNSVNLKL 64
Db 258 PALPTQPAEEAARKREVRMLKMKREAAECRRKKKEYVKCLNRAVAVLEN-QNKTLIEEL 315

RESULT 14
S22298
cAMP response element-binding protein CREB-delta - human
C:Species: Homo sapiens (man)
C>Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C:Accession: S22298
R:Short, M.L.; Manohar, C.F.; Furtado, M.R.; Ghadge, G.D.; Wolinsky, S.M.; Thimmapaya, B
Nucleic Acids Res. 19, 4290, 1991
A:Title: Nucleotide and derived amino-acid sequences of the CRE-binding proteins from ra
A:Reference number: S22298; MUID:91334144; PMID:1831258
A:Accession: S22298
A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: mRNA
A:Residues: 1-327 <SHO>
A:Cross-references: UNIPROT:PI6220; UNIPARC:UPI000016A784; EMBL:X60003; NID:g30493; PIDN:
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1991
C:Keywords: DNA binding; transcription regulation
F:261-304/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 25.9%; Score 88.5; DB 2; Length 327;
Best Local Similarity 33.9%; Pred. No. 0.44;
Matches 20; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

QY 6 PRKRAKTEEKQRRRIILNRRAAHQSRKKRLHLOYLERKCSLLENLLNSVNLKL 64
Db 258 PALPTQPAEEAARKREVRMLKMKREAAECRRKKKEYVKCLNRAVAVLEN-QNKTLIEEL 315

RESULT 15
A35663
cAMP response element-binding protein, hepatic - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 28-Sep-1990 #sequence_revision 28-Sep-1990 #text_change 15-Mar-2004
C:Accession: A35663
R:Quinn, P.G.; Granner, D.K.
Mol. Cell. Biol. 10, 3357-3364, 1990
A:Title: Cyclic AMP-dependent protein kinase regulates transcription of the phosphoenolpy
A:Reference number: A35663; MUID:90287125; PMID:2141384
A:Accession: A35663
A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A:Molecule type: mRNA
A:Residues: 1-328 <QUI>
A:Cross-references: UNIPARC:UPI000017A525
C:Keywords: liver
F:262-305/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 25.9%; Score 88.5; DB 2; Length 328;
Best Local Similarity 33.9%; Pred. No. 0.44;
Matches 20; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

QY 6 PRKRAKTEEKQRRRIILNRRAAHQSRKKRLHLOYLERKCSLLENLLNSVNLKL 64
Db 259 PALPTQPAEEAARKREVRMLKMKREAAECRRKKKEYVKCLNRAVAVLEN-QNKTLIEEL 316

Search completed: November 23, 2005, 03:26:04
Job time : 4.7788 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM protein - protein search, using sw model

Run on: November 22, 2005, 15:37:57 ; Search time 33.5 Seconds
(without alignments)
157.948 Million cell updates/sec

Title: US-10-663-450-6_COPY_53_116

Perfect score: 324

Sequence: 1 KPAKKRKSQGQELPVKTN.....RAAAQTSRRERKLEMEKLES 64

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5 COMB.pap:*
- 2: /cgn2_6/ptodata/1/iaa/6 COMB.pap:*
- 3: /cgn2_6/ptodata/1/iaa/H COMB.pap:*
- 4: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pap:*
- 5: /cgn2_6/ptodata/1/iaa/RE COMB.pap:*
- 6: /cgn2_6/ptodata/1/iaa/backfilee1.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	277	85.5	409	2	US-09-533-029-104
2	166	51.2	248	2	US-09-248-796A-18840
3	100	30.9	143	2	US-09-840-211A-1930
4	91.5	28.2	102	2	US-09-840-211A-1046
5	91	28.1	264	2	US-09-949-016-10789
6	87.5	27.0	260	2	US-09-538-092-950
7	86	26.5	667	2	US-09-248-796A-14492
8	84.5	26.1	521	1	US-08-721-684C-2
9	84.5	26.1	521	1	US-09-005-970-2
10	84.5	26.1	521	2	US-09-407-715-2
11	83	25.6	23	2	US-09-831-642-72
12	80	24.7	212	2	US-09-489-039A-13074
13	79	24.4	63	2	US-09-107-433-3674
14	78	24.1	611	2	US-09-216-393B-81
15	77.5	23.9	1601	2	US-09-345-473E-40
16	77.5	23.9	1601	2	US-09-862-027-40
17	76.5	23.6	306	2	US-09-758-759-115
18	76.5	23.6	395	2	US-09-247-155-113
19	76.5	23.6	395	2	US-09-513-999C-14
20	76.5	23.6	395	2	US-09-471-276-14
21	76.5	23.6	395	2	US-09-503-190-113
22	75.5	23.3	351	2	US-09-870-089B-2
23	75.5	23.3	362	2	US-09-949-016-10923
24	75.5	23.3	501	2	US-09-949-016-11281
25	75.5	23.3	931	2	US-09-949-016-11281
26	74.5	23.0	338	1	US-08-218-686-2
27	74.5	23.0	338	2	US-08-460-242-2

28	74.5	23.0	452	2	US-09-949-016-7289	Sequence 7289, Ap
29	74	22.8	616	2	US-09-873-404-4	Sequence 4, Appli
30	74	22.8	616	2	US-10-243-735-4	Sequence 4, Appli
31	73	22.5	551	2	US-08-796-899-29	Sequence 29, Appl
32	73	22.5	1637	2	US-09-718-692-2	Sequence 2, Appli
33	73	22.5	1637	2	US-09-718-852-2	Sequence 2, Appli
34	73	22.5	1637	2	US-09-718-815-2	Sequence 2, Appli
35	72.5	22.4	371	2	US-09-148-545-259	Sequence 259, App
36	72.5	22.4	371	2	US-09-621-011-259	Sequence 259, App
37	72	22.2	645	2	US-09-949-016-11022	Sequence 11022, A
38	72	22.2	650	2	US-09-487-558B-430	Sequence 430, Ap
39	72	22.2	700	2	US-09-831-642-34	Sequence 34, Appl
40	71	21.9	325	2	US-09-267-031-14	Sequence 14, Appl
41	71	21.9	802	2	US-09-823-240A-2	Sequence 2, Appli
42	70.5	21.8	118	2	US-09-134-001C-2856	Sequence 2856, Ap
43	70.5	21.8	182	2	US-09-640-211A-800	Sequence 800, App
44	70.5	21.8	307	2	US-09-267-031-6	Sequence 6, Appli
45	70	21.6	472	2	US-09-520-781-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1
US-09-533-029-104
; Sequence 104, Application US/09533029
; Patent No. 6664446
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Broun, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Keddle, James
; APPLICANT: Pineda, Omaira
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Reuber, Lynne
; TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES
; FILE REFERENCE: MBI-010
; CURRENT APPLICATION NUMBER: US/09/533,029
; CURRENT FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 60/125,814
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 104
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1034
US-09-533-029-104

Query Match 85.5%; Score 277; DB 2; Length 409;
Best Local Similarity 84.1%; Pred. No. 7.8e-25;
Matches 53; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY	1	KPAKKRKSQGQELPVKTNLPKRAKTEDEKEQRIERVLRNRAAAQTSRRERKLEME 60
DB	67	KKPIKKRKSQGQELPVKTNLPKRAKTDQDEKEQRIERVLRNRAAAQTSRRERKQVE 126
QY	61	KLE 63
DB	127	ALE 129

RESULT 2
US-09-248-796A-18840
; Sequence 18840, Application US/09248796A

```

; Patent No. 6747137
;
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18840
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (221)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown
; US-09-248-796A-18840

```

Query Match 51.2%; Score 166; DB 2; Length 248;
Best Local Similarity 68.1%; Pred. No. 6.2e-12;
Matches 32; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

Qy 18 KTNLP RKRAKTEDEKEQRRIERVLRNRAAAQTSPRKRLEMEKLES 64
 | : ||||| |:|||||:|||||:|||||:|||||:|||||:
Dd 72 KSTLPPRKAKTOEKEORKIERILRNRRRAAHASREKKRHVEYLEN 118

RESULT 3
US-09-640-211A-1930
; Sequence 1930, Application US/09640211A

```

; Patent No. 6833446
;
; GENERAL INFORMATION:
;
; APPLICANT: Wood, Marion
; APPLICANT: Shenk, Michael A.
; APPLICANT: McGrath, Annette
; APPLICANT: Glenn, Matthew
;
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Modification of Gene Transcription
; FILE REFERENCE: 11000.1021CIU
;
; CURRENT APPLICATION NUMBER: US/09/640,211A
; CURRENT FILING DATE: 2000-08-16
;

```

```

; NUMBER OF SEQ ID NOS: 2368
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1930

```

```
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-640-211A-1930
```

Query Match 30.9%; Score 100; DB 2; Length 143;
Best Local Similarity 47.6%; Pred. No. 0.00022;
Matches 20: Conservative 8; Mismatches 14; Indels

QY 22 PPRKRAKTEDEKEQORRIERVLRNRAAAQTSRERKRLEMEKLE 63
| | | | | : | | | | | | | | | | : |
dP 35 PPRKBRADLNAEORREARHNR IAAONS RDKRKAQFTYME 76

RESULT 4
US-09-640-211A-1046
: Sequence 1046. Application US/09640211A

```

; PATENT NO. 6833446
;
; GENERAL INFORMATION:
;
; APPLICANT: Wood, Marion
; APPLICANT: Sherk, Michael A.
; APPLICANT: McGrath, Mnette
; APPLICANT: Glenn, Matthew
;
; TITLE OF INVENTION: Compositions and Methods for the

```

```

; TITLE OF INVENTION: Modification of Gene Transcription
; FILE REFERENCE: 11000.1021CIU
; CURRENT APPLICATION NUMBER: US/09/640,211A
; CURRENT FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 2368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1046
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-640-211A-1046

```

Query Match	28.2%;	Score 91.5;	DB 2;	Length 102;
Best Local Similarity	40.0%;	Pred. No. 0.0015;		
Matches 20: Conservative	11: Mismatches 18;	Indels 1;	Caps 1;	

Qy 14 LPVPKTNLPPrKRakTEDEkQRRIERVLNRAAQTsrEKRLMEKLE 63
:
:
:
Db 31 VVYP-AOANPRKPRSDLSAEKREARAHENRIAONSRDKRKOOFtSLE 79
:
:
:

RESULT 5

US-09-949-016-10789
; Sequence 10789, Application US/09949016
; Patent No. 6812339

; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.

```

/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CLO01307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 10789
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Human
UIS-09-949-016-10789

```

Query Match	28.1%	Score 91	DB 2	Length 264
Best Local Similarity	40.3%	Pred. No. 0.005		
Matches	27	Conservative	9	Mismatches 17
		Indels	14	Gaps 4

Qy 11 GQLP--VP-----KTNLP-PRKRAK-TEDEKEQRRIERVLRNPAQAOTSRRKR 56
||| ||| ||| : | : : ||| : ||| : ||| :
Db 36 GAALPLWYPAORGASPEAAAGGLPOARKRRLTHLSPEEKARERKLKNRVAAOATARDKK 95

QY 57 LEMKLE 63
| : ||
pB 96 ARMSELE 10

RESULT 6

US-09-538-092-950
; Sequence 950, Application US/09538092
; Patent No. 6753314

```

; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield

```

```

; FILE REFERENCE: 15966-542
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
;
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
;
; CURRENT APPLICATION NUMBER: 60/178,965
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
;

```

```
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CufPatSeqFormatter Version 0.9
; SEQ ID NO 950
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P17861
US-09-538-092-950

Query Match      27.0%; Score 87.5; DB 2; Length 260;
Best Local Similarity 46.3%; Pred. No. 0.013;
Matches 19; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

QY 24 RKRK-TEDEKORRIEVLNRRAAAQTSTRKRLEMEKLE 63
Db 58 RKRQRLTHLSPESKALRRKLKRVAAQTARDRKARMSLE 98

RESULT 7
US-09-248-796A-14492
; Sequence 14492, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14492
; LENGTH: 667
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-14492

Query Match      26.5%; Score 86; DB 2; Length 667;
Best Local Similarity 33.3%; Pred. No. 0.056;
Matches 25; Conservative 13; Mismatches 25; Indels 12; Gaps 2;

QY 2 KPAKKGKSGQLPVPKTNLPK-----RAKTEDEKQORRI-----ERVLRNRAAAQ 49
Db 170 KPAKKGKAPAGLAALKQLKLEEQRLLEEEQRLLEEEERLAAEEAKKEAARA 229

QY 50 TSRERKRLEMEKLES 64
Db 230 AKKERLRKKEQKA 244

RESULT 8
US-08-721-684C-2
; Sequence 2, Application US/08721684C
; Patent No. 5854016
; GENERAL INFORMATION:
; APPLICANT: Keegan, Kathleen S.
; TITLE OF INVENTION: No. 5854016el CREBA Isoform
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drvie, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/005,970
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/721,684
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/33487
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 521 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-005-970-2

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/721,684C
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/721,684
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/33487
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 521 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-721-684C-2

Query Match      26.1%; Score 84.5; DB 1; Length 521;
Best Local Similarity 35.6%; Pred. No. 0.064;
Matches 21; Conservative 13; Mismatches 20; Indels 5; Gaps 2;

QY 5 KKRKSGQBLPVPKTNLPKRAKTEDEKQORRIEVLNRRAAAQTSTRKRLEMEKLE 63
Db 270 EKRTLVAEGYPIP-TKLPIKTK---SEKALKIRKIKKIKSAQESRRKKEYMDSLE 323

RESULT 9
US-09-005-970-2
; Sequence 2, Application US/09005970
; Patent No. 5959079
; GENERAL INFORMATION:
; APPLICANT: Keegan, Kathleen S.
; TITLE OF INVENTION: No. 5959079el CREBA Isoform
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drvie, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/005,970
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/721,684
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/33487
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 521 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-005-970-2
```


US-09-345-473E-40 ; Sequence 40, Application US/09345473E
 ; Patent No. 6558903
 ; GENERAL INFORMATION:
 ; APPLICANT: Hodge, Martin

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2005, 15:33:00 ; Search time 147.5 Seconds
(without alignments)
306.128 Million cell updates/sec

Title: US-10-663-450-6_COPY_53_116
Perfect score: 324
Sequence: 1 KKPAAKKKSWGQELPVPTKN.....RAAAQTSRRKRLEMEKLES 64

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	324	100.0	347	2	Q5AQN3_EMENI	Q5aqn3 aspergillus
2	324	100.0	350	2	Q8TFU8_EMENI	Q8tfu8 emericella
3	317	97.8	342	2	Q6W8X2_ASPNG	Q6w8x2 aspergillus
4	315	97.2	433	2	Q4WEY8_ASPFU	Q4wey8 aspergillus
5	283	87.3	451	2	Q8TFP3_TRIRE	Q8tfp3 trichoderma
6	282	87.0	430	2	Q7SHF0_NEUCR	Q7shf0 neurospora
7	265	81.8	556	2	Q51KW8_MAGGR	Q51kw8 magnaporthe
8	250	77.2	429	2	Q4HTY5_GIBZE	Q4hty5 gibberella
9	178.5	55.1	299	2	Q6CEV1_YARLI	Q6cevi yarrowia li
10	174.5	53.9	273	2	Q6CKQ1_KLULA	Q6ckq1 kluyveromyc
11	171.5	52.9	230	1	HAC1_YEAST	P41546 saccharomyc
12	171	52.8	260	2	Q6B0C2_DEBHA	Q6bgc2 debaryomyc
13	166	51.2	228	2	Q75BQ5_ASHGO	Q75bq5 ashbya goss
14	166	51.2	357	2	Q5AA52_CANAL	Q5aa52 candida alb
15	144	44.4	329	2	Q6FLY3_CANGA	Q6fly3 candida gla
16	117	36.1	336	2	Q5VR11_ORYSA	Q5vr11 oryza sativ
17	116	35.8	168	1	HY5_ARATH	Q24646 arabidopsis
18	109	33.6	69	2	Q58L60_BRACH	Q58l60 brassica ca
19	109	33.6	109	2	Q84XX8_ERARP	Q84xx8 brassica ra
20	108	33.3	176	2	Q69XK6_ORYSA	Q69xk6 oryza sativ
21	105.5	32.6	141	2	Q677A7_9ASPA	Q677a7 hyacinthus
22	104.5	32.3	252	2	Q6NX18_XENTR	Q6nx18 xenopus tro
23	103.5	31.9	158	1	HY5_LYCES	Q9sm50 lycopersico
24	102.5	31.6	350	2	Q90Zr7_XENLA	Q90zr7 xenopus lae
25	102.5	31.6	396	2	Q7ZYC2_XENLA	Q7zyc2 xenopus lae
26	100	30.9	208	2	Q5DFK2_SCHJA	Q5dfk2 schistosom
27	100	30.9	321	2	Q8GRY7_LOTJA	Q8gry7 lotus japon
28	99	30.6	188	2	Q6ZHT8_ORYSA	Q6zht8 oryza sativ
29	98	30.2	120	2	Q682B6_ARATH	Q682b6 arabidopsis
30	98	30.2	149	1	HYH_ARATH	Q8w191 arabidopsis
31	97	29.9	322	2	Q39896_SOYEN	Q39896 glycine max

32	97	29.9	326	2	Q39895_SOYEN	Q39895 glycine max
33	95	29.3	703	2	Q4WC74_ASPFU	Q4wc74 aspergillus
34	94.5	29.2	309	2	Q4H2M2_CIOIN	Q4h2m2 ciona intes
35	94	29.0	322	2	O04234_VICFA	O04234 vicia faba
36	94	29.0	627	2	Q5BD44_EMENI	Q5bd44 aspergillus
37	92	28.4	263	2	Q8UVQ5_BRARE	Q8uvq5 brachydanio
38	92	28.4	263	2	Q90X27_BRARE	Q90x27 brachydanio
39	92	28.4	383	2	Q8QHJ5_BRARE	Q8qhj5 brachydanio
40	92	28.4	383	2	Q90XD3_BRARE	Q90xd3 brachydanio
41	92	28.4	646	2	Q6AU90_ORYSA	Q6au90 oryza sativ
42	91.5	28.2	260	2	Q6EZA7_OREMO	Q6eza7 oreochromis
43	91.5	28.2	260	2	Q6EZA8_OREMO	Q6eza8 oreochromis
44	91	28.1	261	1	XPB1_HUMAN	XPB1 homo sapien
45	90.5	27.9	176	2	Q98RX1_GUITH	Q98rx1 guillardia

ALIGNMENTS

RESULT 1
Q5AQN3_EMENI PRELIMINARY; PRT; 347 AA.
AC Q5AQN3;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Hypothetical protein.
GN ORFNames=AN9397.2;
OS Aspergillus nidulans FGSC A4.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=227321;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FGSC A4;
RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavskiy L.,
RA Boukhgater B., Butler J., Calvo S.E., Camarata J., Chang J.,
RA Choepel Y., Collymore A., Cooke P., Corum B., DeArelano K.,
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
RA Erickson J., Faro S., Ferreira P., FitzGerald M., Gage D., Galagan J.,
RA Gardyna S., Gnarr S., Graham L., Grand-Pierre N., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,
RA Jaffe D., Johnson R., Jones C., Kanat M., Kanat A., Karatas A.,
RA Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
RA Ma L.-J., Mabbitt R., Maclean C., Macdonald P., Major J., Manning J.,
RA Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,
RA Mihova T., Mlenga V., Murphy T., Naylor J., Nguyen C., Nicol R.,
RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neil D.,
RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,
RA Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,
RA Roman J., Schauer S., Schupback R., Seaman S., Severy P., Smirnov S.,
RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,
RA Talamas J., Tesfaye S., Theodore J., Topham K., Travers M.,
RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
RA Lander E.;
RT "Genome Sequence of Aspergillus nidulans";
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
EMBL: AACD01000172; EAA66464.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
SQ DNA-binding; Hypothetical protein; Nuclear protein.
KW SEQUENCE 347 AA; 37841 MW; A7CAAFEE761B3E1 CRC64;

Query Match 100.0%; Score 324; DB 2; Length 347;
Best Local Similarity 100.0%; Pred. No. 7.2e-22;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KKPAAKKKSWGQELPVPTKNLPKRAKTEDEKEQRRIERVLNRNRAAAQTSRRKRLEME 60
DB 54 KKPAAKKKSWGQELPVPTKNLPKRAKTEDEKEQRRIERVLNRNRAAAQTSRRKRLEME 113

QY 61 KLES 64
 ||||
 Db 114 KLES 117

RESULT 2

Q8TFUB EMENI
 ID Q8TFUB_EMENI PRELIMINARY; PRT; 350 AA.
 AC Q8TFUB8;
 DT 01-JUN-2002 (TREMELrel. 21, Created)
 DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)
 DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
 DE Transcription factor.
 GN Name-hacA;
 OS Emericella nidulans (Aspergillus nidulans).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eutotiales; Trichocomaceae; Emericella.
 OX NCBI_TaxID=162425;
 RN [1]

NUCLEOTIDE SEQUENCE.

RA Salotheimo M.L.A., Valkonen M., Penttilae M.E.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ413273; CAC88375.1; -; Genomic_DNA.
 DR HSSP; P05412; 1JNM.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0003555; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR011700; bZIP 2.
 DR InterPro; IPR004827; TF_bZIP.
 DR Pfam; PF07716; bZIP 2; 1.
 DR SMART; SM00338; BRLZ; 1.
 DR PROSITE; PS0217; bZIP; 1.
 DR PROSITE; PS00036; bZIP_BASIC; UNKNOWN_1.
 KW DNA-binding; Nuclear protein.
 SQ SEQUENCE 350 AA; 38192 MW; 4F78340096064F44 CRC64;

Query Match 100.0%; Score 324; DB 2; Length 350;
 Best Local Similarity 100.0%; Pred. No. 7.3e-22;
 Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKPAAKRSKSGQELPVPKTNLPKRKRAKTEDEKEQRRIERVLRNRAAAQTSRRKRELEME 60
 |||||
 Db 54 KKPAAKRSKSGQELPVPKTNLPKRKRAKTEDEKEQRRIERVLRNRAAAQTSRRKRELEME 113
 |||||
 QY 61 KLES 64
 ||||
 Db 114 KLES 117

RESULT 3

Q6W8X2 ASPNG
 ID Q6W8X2_ASPNG PRELIMINARY; PRT; 342 AA.
 AC Q6W8X2;
 DT 05-JUL-2004 (TREMELrel. 27, Created)
 DT 05-JUL-2004 (TREMELrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMELrel. 27, Last annotation update)
 DE Transcription factor HACa.
 GN Name-hacA;
 OS Aspergillus niger.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eutotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=5061;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Mulder H.J.;
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY303684; AAQ73495.1; -; Genomic_DNA.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0003555; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR011700; bZIP 2.
 DR InterPro; IPR004827; TF_bZIP.

DR Pfam; PF07716; bZIP 2; 1.
 DR SMART; SM00338; BRLZ; 1.
 DR PROSITE; PS0217; bZIP; 1.
 DR PROSITE; PS00036; bZIP_BASIC; UNKNOWN_1.
 KW DNA-binding; Nuclear protein.
 SQ SEQUENCE 342 AA; 37148 MW; 623B3941A55C9C05 CRC64;

Query Match 97.8%; Score 317; DB 2; Length 342;
 Best Local Similarity 96.9%; Pred. No. 3.1e-21;
 Matches 62; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKPAAKRSKSGQELPVPKTNLPKRKRAKTEDEKEQRRIERVLRNRAAAQTSRRKRELEME 60
 |||||
 Db 47 KKPAAKRSKSGQELPVPKTNLPKRKRAKTEDEKEQRRIERVLRNRAAAQTSRRKRELEME 106
 |||||
 QY 61 KLES 64
 ||||
 Db 107 KLEN 110

RESULT 4

Q4WEY8 ASPFU
 ID Q4WEY8_ASPFU PRELIMINARY; PRT; 433 AA.
 AC Q4WEY8;
 DT 13-SEP-2005 (TREMELrel. 31, Created)
 DT 13-SEP-2005 (TREMELrel. 31, Last sequence update)
 DT 13-SEP-2005 (TREMELrel. 31, Last annotation update)
 DE bZIP transcription factor (HacA), putative.
 GN ORFNames=Afu3G04070;
 OS Aspergillus fumigatus Af293.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eutotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=330879;
 RN [1]

NUCLEOTIDE SEQUENCE.

RC STRAIN-Af293;
 RA Nierman W., Pain A., Anderson M.J., Wortman J., Kim H.Stanley.,
 RA Arroya J., Berrihan M., Abe K., Archer D.B., Bermejo C., Bennett J.,
 RA Bowyer P., Chen D., Collins M., Coulson R., Davies R., Dyer P.S.,
 RA Farman M., Pedorova N., Pedorova N., Feldblyum T.V., Fischer R.,
 RA Fosker N., Fraser A., Garcia J.L., Garcia M.J., Goble A.,
 RA Goldman G.H., Gomi K., Griffith-Jones S., Gwilliam R., Haas B.,
 RA Haas H., Harris D., Horiuchi H., Huang J., Humphrey S., Jimenez J.,
 RA Keller N., Khouri H., Kitamoto K., Kobayashi T., Kulkarni R.,
 RA Kumagai T., Lafton A., Latge J.-P., Li W., Lord A., Lu C.,
 RA Majoros W.H., May G.S., Miller B.B., Mohamoud Y., Molina M., Monod M.,
 RA Mouyna I., Mulligan S., Murphy L., O'Neil S., Paulsen I.,
 RA Penalba M.A., Pettea M., Price C., Pritchard B.L., Quail M.A.,
 RA Rabinowitsch E., Rawlins N., Rajandream M.-A., Reichard U.,
 RA Renauld H., Robson G.D., Rodriguez de Cordoba S., Rodriguez-Pena J.M.,
 RA Ronning C.M., Rutter S., Salzberg S.L., Sanchez M.,
 RA Sanchez-Ferrero J.C., Saunders D., Seeger K., Squares R., Squares S.,
 RA Takeuchi M., Tekala F., Turner G., Vazquez de Aldana C.R., Weidman J.,
 RA White O., Woodward J., Yu J.-H., Fraser C., Galagan J.E., Asai K.,
 RA Machida M., Hall N., Barrell B., Denning D.W.;
 RT "Genomic sequence of the pathogenic and allergenic filamentous fungus
 Aspergillus fumigatus."
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: the sequence shown here is derived from an
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 preliminary data.
 CC EMBL; AAHF01000010; EAL86689.1; -; Genomic_DNA.
 DR InterPro; IPR011700; bZIP 2.
 DR InterPro; IPR004827; TF_bZIP.
 DR Pfam; PF07716; bZIP 2; 1.
 DR SMART; SM00338; BRLZ; 1.
 DR PROSITE; PS0217; bZIP; 1.
 DR PROSITE; PS00036; bZIP_BASIC; UNKNOWN_1.
 KW DNA-binding; Nuclear protein.
 SQ SEQUENCE 433 AA; 48908 MW; 3742DD0B43E05C74 CRC64;

Query Match 97.2%; Score 315; DB 2; Length 433;
 Best Local Similarity 95.3%; Pred. No. 6e-21;

RA Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamvassilis M., Mauceli E., Biele C., Rudd S., Frishman D.,
RA Krystofova S., Rasmussen C., Metzenberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Omani S.A.,
RA Desouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.,
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
EL Nature 0-0-0(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AA001000004; EAA36251.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; SZIP_2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07716; bZIP 2; I.
DR PROSITE; PS02117; bZIP; 1.
SQ SEQUENCE 430 AA; 45599 MW; 0EE0657CFA6160DA CRC64;

Query Match 87.0%; Score 282; DB 2; Length 430;
Best Local Similarity 87.3%; Pred. No. 6.5e-18;
Matches 55; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 KPAKKKSGOELPVKPTNLPPRKAKTDEKEORRIERVLNRRAAOTSRERKLEME 60
DB 88 KKPVKRKSGVQLPEPTNLPPKGAKEDEKEORRVERVLNRRAAOSRRKRLEVE 147

QY 61 KLE 63
DB 148 GLR 150

RESULT 7
Q51KW8 MAGGR
ID Q51KW8_MAGGR PRELIMINARY; PRT; 556 AA.
AC Q51KW8;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORENAMES-MG09010.4;
OS Magnaporthe oryzae 70-15.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
NCBI_TaxID=242507;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Birren B., Nussbaum C., Abebe A., Abouelleil A., Adekoya E.,
RA Ait-zahra M., Allen T., An P., Anderson M., Anderson S.,
RA Arachihi H., Armbruster J., Bachantsang P., Baldwin J., Barry A.,
RA Bayul T., Blithsteyn B., Bloom T., Blye J., Boguelavskiy L.,
RA Borowsky M., Bouhgalter B., Brunache A., Butler J., Calixte N.,
RA Calvo S., Camarata J., Campo K., Chang J., Cheshatsang Y., Citroen M.,
RA Collimore A., Considine T., Cook A., Cooke P., Corum B., Cuomo C.,
RA David R., Dawoe T., Degray S., Dodge S., Dooley K., Dorje P.,
RA Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T., Engels R.,
RA Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,
RA Fitzgerald M., Foley K., Gage D., Galegan J., Gearin G., Gnerre S.,
RA Gnirke A., Goyette A., Graham J., Grandbois E., Gyaltzen K., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,
RA Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,
RA Jaffe D., Jones C., Kamal M., Kamat A., Kamvassilis M., Karlsson E.,
RA Kells C., Kieu A., Kinser P., Kodira C., Kulbokas E., Labutti K.,
RA Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,
RA Lindblad-toh K., Liu X., Lokyitsang T., Lokyitsang Y., Lucien O.,
RA Lui A., Ma L.J., Mabbitt R., MacDonald J.J., Maclean C., Major J.,

```
RA Manning J., Marabella R., Maru K., Matthews C., Mauceli E.,
RA McCarthy M., McDonough S., Mcghee T., Meldrim J., Meneus L.,
RA Mesirov J., Mihalov A., Mihova T., Mikkelsen T., Mlenga V., Moru K.,
RA Mozes J., Mulrain L., Munson G., Naylor J., Newes C., Nguyen C.,
RA Nguyen N., Nguyen T., Nicol R., Nielsen C., Nizzari M., Norbu C.,
RA Norbu N., O'donnell P., Okoawo O., O'leary S., Omotosho B.,
RA O'neill K., Osman S., Parker S., Perrin D., Phunkhang P., Pigani B.,
RA Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond C.,
RA Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,
RA Rutnan M., Schubach R., Seaman C., Settippalli S., Sharpe T.,
RA Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,
RA Spencer B., Stalker J., Stange-Thomann N., Stavropoulos S.,
RA Stenson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,
RA Tensing P., Tesfaye S., Theodore J., Thoulutsang Y., Topham K.,
RA Towey S., Teamla T., Tsomo N., Vallee D., Vassiliev H.,
RA Venkataraman V., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,
RA Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,
RA Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembek L.,
RA Zimmer A., Zody M., Lander E.;
RT "The genome sequence of Magnaporthe grisea.";
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Dean R., Mitchell T., Brown D., Pan H., Thon M.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Zhu H., Blackmon B.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
preliminary data.
DR EMBL; AACU01001606; EAA7880.1; -; Genomic_DNA.
DR InterPro; IPR011700; bZIP 2.
DR InterPro; IPR004827; TF bZIP.
DR Pfam; PF07716; bZIP 2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS50217; bZIP; 1.
KW DNA-binding; Hypothetical protein; Nuclear protein.
SQ SEQUENCE 556 AA; 60798 MW; CFDC4C8D2DC4963C CRC64;

Query Match 81.8%; Score 265; DB 2; Length 556;
Best Local Similarity 82.5%; Pred. No. 3.1e-16;
Matches 52; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 KKPAAKRSWGQELPVPKTNLPKRAKTEDEKEQRIERVLNRRAAAQTSRERKLEME 60
: ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 88 EKTKRKRSWGQVLPPEKTNLPKRAKTEDEKEQRIERVLNRRAAAQTSRERKQVE 147
QY 61 KLE 63
DB 148 ALE 150

RESULT 8
Q4HTT5_GIBZE
ID Q4HTT5_GIBZE PRELIMINARY; PRT; 429 AA.
AC Q4HTT5;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=FGL1623.1;
OS Gibberella zeae PH-1.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
ON NCBI_TaxID=229533;
RX STRAIN=PH-1;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PH-1;
RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
```

```
RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavskiy L.,
RA Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
RA Choepl Y., Collymore A., Cook A., Cooke P., Corum B., Dearellano K.,
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
RA Erickson J., Faro S., Ferreira P., Fitzgerald M., Gage D., Galagan J.,
RA Gardyna S., Gnerre S., Graham L., Grand-Pierre N., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,
RA Jaffe D., Johnson R., Jones C., Kamal M., Kamat A., Karatas A.,
RA Kellis C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
RA Ma L.-J., Mabbitt R., MacLean C., MacDonald P., Major J., Manning J.,
RA Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,
RA Mihova T., Mlenga V., Murphy T., Naylor J., Nguyen C., Nicol R.,
RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neil D.,
RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,
RA Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,
RA Roman J., Schauer S., Schuback R., Seaman S., Severy P., Smirnov S.,
RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,
RA Talamas J., Tesfaye S., Theodore J., Topham K., Travers M.,
RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
RA Lander E.;
RT "Fusarium graminearum genome sequence.";
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
preliminary data.
DR EMBL; AACM01000497; EAA78735.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 429 AA; 46630 MW; 04CDE40BD392901 CRC64;

Query Match 77.2%; Score 250; DB 2; Length 429;
Best Local Similarity 65.4%; Pred. No. 5.7e-15;
Matches 53; Conservative 3; Mismatches 7; Indels 18; Gaps 1;

QY 1 KKPAAKRSWGQELPVPKTNLP-----RKQAKTEDEKEQRIERVL 42
DB 83 KTKTKRKRSWGQVLPPEKTNLPYVDTLRHVVDSTDTLNRKRAKTEDEKEQRIERVL 142
QY 43 RNRAAQSRRERKLEMEKLE 63
DB 143 RNRAAQSRRERKQVEALE 163

RESULT 9
Q6CEV1_YARL1
ID Q6CEV1_YARL1 PRELIMINARY; PRT; 299 AA.
AC Q6CEV1;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Yarrowia lipolytica chromosome B of strain CLIB99 of Yarrowia
DE lipolytica.
GN OrderedLocNames=YALI0B12716g;
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
ON NCBI_TaxID=4952;
RX STRAIN=CLIB 122 / E 150;
RC PubMed=15229592; DOI=10.1038/nature02579;
RA Dujon B., Sherman D., Fischer G., Duxrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
RA Boissarie A., Boyer E., Cattolico L., Confanioli F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Jauniaux N., Joyet P., Kachouri R.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,
RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic Y., Boiotin-Fukuhara M., Thierry A.,
```

Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J., Wincker P., Souciet J.-L.; "Genome evolution in yeasts."; Nature 430:35-44(2004).

EML; CR382128; CAG83062.1; -: Genomic_DNA.

GO; GO:0005634; C:nucleus; IEA.

GO; GO:0003677; F:DNA binding; IEA.

GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

InterPro; IPR004827; TF bZIP.

Pfam; PF07116; bZIP_2; 1.

SMART; SM00338; BRZ; 1.

PROSITE; PS02117; bZIP; 1.

DR PROSITE; PS00036; bZIP_BASIC; UNKNOWN_1.

KW Complete proteome; DNA-binding; Nuclear protein.

SQ SEQUENCE 299 AA; 32817 MW; P526110CFB23ABB2 CRC64;

Query Match 55.1%; Score 178.5; DB 2; Length 299;

Best Local Similarity 65.5%; Pred. No. 1.5e+08;

Matches 38; Conservative 6; Mismatches 13; Indels 1; Gaps 1;

6 KRKSWGQLPVPTNLPKRRAKTEDSKQRIRIVLRNRAAQTSTRRKLEMEKL 63
|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||

33 KKKK-DUTLPDAGALPDKRKATENEKEQRRIWRNQAHASREKKRHLEDLE 89
|||||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT 10

Q6CKQ1_KJULJA

ID Q6CKQ1_KJULA PRELIMINARY; PRT; 273 AA.

AC Q6CKQ1;

DT 25-OCT-2004 (TrEMBLrel. 28, Created)

DT 25-OCT-2004 (TrEMBurel. 28, Last sequence update)

DT 25-OCT-2004 (TrEMBurel. 28, Last annotation update)

DE Kluyveromyces lactis strain NRRL Y-1140 chromosome F of strain NRRL Y-

DE 1140 of Kluyveromyces lactis.

GN OrderedLocustNames=KLUAF08976g;

OS Kluyveromyces lactis (Yeast)

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

CC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.

OX NCBI_TaxID=28985;

RN [1]

NN NUCLEOTIDE SEQUENCE [Large scale genomic DNA].

RP STRAIN=CBS 2359 /IFO 1267 /NRRL Y-1140 /WM37;

RC PubMed=1529592; DOI=10.1038/nature02579;

RA Dujon B., Sherman D., Fischer G., Durans P., Casaregola S., Lafontaine I., de Montigny J., Marck C., Neugeuse C., Talla E., Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V., Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykassen C., Boissame A., Boyer J., Cattolic L., Confanieri F., de Daruvar A., Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A., Hanxaye F., Hennerquin C., Jauniaux N., Joyet P., Kachouri R., Kerrest A., Koszul R., Lenaire M., Lesur I., Ma L., Muller H., Nicaut J.-M., Nikolaiki M., Orlas S., Oxler-Kalogeropoulous O., Pellens S., Potier S., Richard G.-F., Straub M.-L., Suleau A., Svennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B., Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A., Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J., Wincker P., Souciet J.-L.; "

"genome evolution in yeasts.";

Nature 430:35-44(2004).

RL ENMBL; CR382126; CAG98196.1; -: Genomic_DNA.

GO; GO:0005634; C:nucleus; IEA.

GO; GO:0003677; F:DNA binding; IEA.

GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

InterPro; IPR011616; bZIP_1.

PFam; PF00170; bZIP_1; 1.

SMART; SM00338; BRZ; 1.

DR PROSITE; PS00036; bZIP_BASIC; 1.

KW Complete proteome; DNA-binding; Nuclear protein.

SQ SEQUENCE 273 AA; 30429 MW; 97DFFD7BF8CACE5 CRC64;

Query Match

53.9%; Score 174.5; DB 2; Length 273;


```

DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR011700; bZIP 2.
DR Pfam: PF07716; bZIP 2; 1.
DR PROSITE: PS00036; bZIP_BASIC; 1.
KW Complete proteome.
SQ SEQUENCE 228 AA; 24914 MW; 605A60D78633F038 CRC64;

Query Match 51.2%; Score 166; DB 2; Length 228;
Best Local Similarity 71.7%; Pred. No. 1.6e-07;
Matches 33; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 18 KTNLPKRAKTEDEKQRIERVLRNRAAAQTSRKRKLEMEKLE 63
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 14 KSTLPKRAKTEDEKQRIERVLRNRAAAQTSRKRKLEMEKLE 59
   |:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 14
Q5AA52 CANAL PRELIMINARY; PRT; 357 AA.
AC Q5AA52;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Hypothetical protein HAC1.
GN Names=HAC1; ORFNames=CaO19.2432, CaO19.9968;
OS Candida albicans SC5314.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=537561;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SC5314;
RX PubMed=15123810; DOI=10.1073/pnas.0401648101;
RA Jones T., Federapfel N.A., Chibana H., Dungan J., Kalman S., Magee P.T.,
RA Magee B.B., Newport K., Thorntonsen Y.R., Agabian N.,
RA Davis R.W., Scherer S.;
RT "The diploid genome sequence of Candida albicans."
RL Proc. Natl. Acad. Sci. U.S.A. 101:7329-7334(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SC5314;
RA Dungan J., Kuo A., Newport G., Lan C.-Y., Iijima C., Adegbola O.,
RA Roberts J., Persson K., Donnelly S., Favorito S., Tsung K.-W.,
RA Jones T., Scherer S., Agabian N.;
RT "Annotation of the Genome of Candida albicans."
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AACQ01000041; EAK99529.1; -; Genomic DNA.
DR EMBL; AACQ01000040; EAK99617.1; -; Genomic DNA.
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
KW DNA-binding, Hypothetical protein; Nuclear protein.
SQ SEQUENCE 357 AA; 40127 MW; 8E1C633A2B4ECDAB CRC64;

Query Match 51.2%; Score 166; DB 2; Length 357;
Best Local Similarity 68.1%; Pred. No. 2.5e-07;
Matches 32; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 18 KTNLPKRAKTEDEKQRIERVLRNRAAAQTSRKRKLEMEKLES 64
   |:|||||:|||||:|||||:|||||:|||||:|||||:
Db 54 KSTLPKRAKTEDEKQRIERVLRNRAAAQTSRKRKLEMEKLES 100
   |:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 15
Q6FLY3 CANGA PRELIMINARY; PRT; 329 AA.
AC Q6FLY3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)

```

```

DE Candida glabrata strain CBS138 chromosome X complete sequence.
GN OrderedLocusNames=CAGL0K12540g;
OS Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 2001 / CBS 138;
RX PubMed=15229592; DOI=10.1038/nature02579;
RA Lafontaine I., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barny S., Blanchin S., Beckerich J.-M., Beyne E., Breykasteen C.,
RA Boisrame A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lueur I., Ma L., Muller H.,
RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Peillen S., Potier S., Richard G.-F., Straub M.-L., Suteau A.,
RA Zenonen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.-L.;
RT "Genome evolution in yeasts."
RL Nature 430:35-44(2004).
DR EMBL; CR380957; CAG61724.1; -; Genomic DNA.
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR011616; bZIP 1.
DR InterPro: IPR004827; TF_bZIP.
DR Pfam: PF00170; bZIP 1; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS00217; bZIP 1.
DR PROSITE; PS00036; bZIP_BASIC; 1.
KW Complete proteome; DNA-binding; Nuclear protein.
SQ SEQUENCE 329 AA; 37162 MW; 469614BF37AF2282 CRC64;

Query Match 44.4%; Score 144; DB 2; Length 329;
Best Local Similarity 63.0%; Pred. No. 2.5e-05;
Matches 29; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

QY 18 KTNLPKRAKTEDEKQRIERVLRNRAAAQTSRKRKLEMEKLE 63
   |:|||||:|||||:|||||:|||||:|||||:|||||:
Db 26 ETWMPKRAKTEDEKQRIERVLRNRAAAQTSRKRKLEMEKLE 71
   |:|||||:|||||:|||||:|||||:|||||:|||||:

Search completed: November 22, 2005, 15:48:42
Job time : 148.5 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2005, 15:34:11 ; Search time 24 Seconds
(without alignments)
256.578 Million cell updates/sec

Title: us-10-663-450-6_COPY_53_116

Perfect score: 324

Sequence: 1 KPAKTKRKSNGQLPVPKTN.....RAAAQTSRRKRLEMEKLES 64

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 80.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	171.5	52.9	238	2 S78571	transcription fact
2	116	35.8	168	2 T05922	bZIP protein HY5 (
3	97	29.9	322	2 T08592	TGACG-motif-bindin
4	97	29.9	326	2 T08591	TGACG-motif bindin
5	94	29.0	322	2 T12093	TGACG-motif bindin
6	90.5	27.9	176	2 B90087	hypothetical prote
7	90.5	27.9	506	2 D84609	hypothetical prote
8	87.5	27.0	260	1 A36299	transcription fact
9	87.5	27.0	266	2 JC7300	tax-responsive ele
10	87.5	27.0	287	2 JC4857	hepatocarcinogenes
11	86	26.5	600	2 T00759	hypothetical prote
12	83	25.6	688	2 T32750	hypothetical prote
13	82.5	25.5	515	2 A42140	box B-binding fact
14	82.5	25.5	516	2 A44494	CAMP-responsive el
15	81	25.0	360	2 T03373	probable G-box bin
16	81	25.0	488	2 S33222	transcription fact
17	81	25.0	483	2 S12741	transcription fact
18	79.5	24.5	486	2 JC4028	activating transcr
19	77.5	23.9	1851	2 T19964	hypothetical prote
20	77	23.8	433	2 JC1230	DNA-binding protei
21	76.5	23.6	242	2 S05453	transcription fact
22	76	23.5	207	2 T40067	hypothetical prote
23	76	23.5	284	2 T24253	hypothetical prote
24	76	23.5	331	2 S33223	transcription fact
25	76	23.5	445	2 T50972	probable zootin (i
26	75.5	23.3	246	2 T12585	Dc3 promoter-bindi
27	75.5	23.3	313	2 A34785	DNA-binding protei
28	75.5	23.3	349	2 A41349	histone-specific t
29	75.5	23.3	349	2 S77570	transcription fact

30	75.5	23.3	351	2 A45377	transcription fact
31	75.5	23.3	358	2 C42026	cyclic AMP respons
32	75.5	23.3	389	1 A39429	CAMP response elem
33	75.5	23.3	448	2 A42026	cyclic AMP respons
34	75.5	23.3	456	2 B42026	cyclic AMP respons
35	75.5	23.3	505	1 S05380	transcription fact
36	75.5	23.3	849	1 S64732	scaffold attachmen
37	75.5	23.3	1359	2 T34036	hypothetical prote
38	75	23.1	381	2 S26812	transcription fact
39	75	23.1	1549	1 A40691	trichohyalin - she
40	74.5	23.0	313	2 S66312	G-box binding fact
41	74.5	23.0	315	2 S20883	G-box-binding fact
42	74.5	23.0	315	2 G85433	G-box-binding fact
43	74.5	23.0	338	1 TVMSFB	transforming prote
44	74.5	23.0	338	2 I53043	transforming prote
45	74.5	23.0	452	2 H96710	hypothetical prote

ALIGNMENTS

RESULT 1

S78571
transcription factor HAC1 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YFL031w
C;Species: Saccharomyces cerevisiae
C;Date: 13-Sep-1998 #sequence revision 13-Sep-1998 #text_change 09-Jul-2004
C;Accession: S78571; S56223; S53578
R;Murakami, Y.
submitted to the Protein Sequence Database, January 1998
A;Reference number: S78570
A;Accession: S78571
A;Molecule type: DNA
A;Residues: 1-238 <MUR>
A;Cross-references: UNIPROT:P41546; UNIPARC:UPI0000157F61; EMBL:D50617; MIPS:YFL031w
A;Note: this is a revision to the sequence from reference S56186
R;Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sasanuma, S.I.;
submitted to the EMBL Data Library, May 1995
A;Description: Analysis of the nucleotide sequence of chromosome VI from Saccharomyces cerevisiae
A;Reference number: S56186
A;Accession: S56223
A;Molecule type: DNA
A;Residues: 1-131, 'RWQTHSATI' <MUR>
A;Cross-references: UNIPARC:UPI0000179860; EMBL:D50617; NID:g836685; PID:g836723; MIPS:YFL031w
A;Note: this sequence has been revised in reference S78570
A;Note: this was believed to be the complete sequence of YFL031w
R;Nojima, H.; Leem, S.H.; Araki, H.; Sakai, A.; Nakashima, N.; Kanaoka, Y.; Ono, Y.
Nucleic Acids Res. 22, 5279-5288, 1994
A;Title: Hac1: a novel yeast bZIP protein binding to the CRE motif is a multicopy suppressor of the
A;Reference number: S53578; MUID:95116316; PMID:7816617
A;Accession: S53578
A;Molecule type: DNA
A;Residues: 1-142, 'RLCRPRYCRFRVGRDPFGAAECURRMYOSRRYLPVTI', 183-220, 'AVITMTKRLQ' <NOJ>
A;Cross-references: UNIPARC:UPI000012C042; GB:D26506; NID:g633122; PIDN:BAA05513.1; PID:J05513.1
C;Genetics:
A;Gene: SGD:HAC1; IRE2; ERN4
A;Cross-references: SGD:S0001863; MIPS:YFL031w
A;Map position: 6L
A;Introns: 221/1
C;Keywords: DNA binding; nucleus; transcription factor

Query Match 52.9%; Score 171.5; DB 2; Length 238;
Best Local Similarity 66.0%; Pred. No. 3.7e-09;
Matches 35; Conservative 8; Mismatches 7; Indels 3; Gaps 1;

QY 14 LPVP---KTNLPKRKTEDEKQRIERVLRNRAAAQTSRRKRLEMEKLE 63
DB 16 LAIPNFKSTLPPRKAKTKESKQRIERIRLNRRRAHQSRKKKTLHLYLE 68

RESULT 2

T50922
bZIP protein HY5 [imported] - Arabidopsis thaliana

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: D84609

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-506 <STO>

A;Cross-references: UNIPROT:Q9SIF0; UNIPARC:UPI00000A51B3; GB:AE002093; NID:G4567209; PMID:10617197

C;Genetics:

A;Gene: At2g22140

A;Map position: 2

Query Match 27.9%; Score 90.5; DB 2; Length 506;

Best Local Similarity 34.8%; Pred. No. 0.29;

Matches 23; Conservative 13; Mismatches 25; Indels 5; Gaps 1;

QY 1 KKPAKKKSGQLPVPKTLPRKRAKTE-----DEKQRRRIERVLRNRAAAQTSSRERK 55

DB 161 EKTRKKKIRTTTTLPVGEALPKQSKEDKTSAMEEKKLRKEQERLEKAASAEERK 220

QY 56 RLEMEK 61

DB 221 RLEMEK 226

RESULT 8

A36299

transcription factor hXBP-1 - human

N;Alternate names: DNA-binding protein TREBS

C;Species: Homo sapiens (man)

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

C;Accession: A36299; S12559; R47490

R;Liou, H.C.; Boothby, M.R.; Finn, P.W.; Davidson, R.; Nabavi, N.; Zeleznik-Le, N.J.; Tjian, R.

Science 247, 1581-1583, 1990

A;Title: A new member of the leucine zipper class of proteins that binds to the HLA DR alpha 1 promoter

A;Reference number: A36299; MUID:90208323; PMID:2321018

A;Accession: A36299

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-260 <LIO>

A;Cross-references: UNIPROT:P17861; UNIPARC:UPI0000031C60; GB:M31627; NID:G184485; PMID:10617197

R;Yoshimura, T.; Fujisawa, J.I.; Yoshida, M.

EMBO J. 9, 2537-2542, 1990

A;Title: Multiple cDNA clones encoding nuclear proteins that bind to the tax-dependent enhancer

A;Reference number: S12559; MUID:90316112; PMID:2196176

A;Accession: S12559

A;Molecule type: DNA

A;Residues: 1-32, 'GQA', 35-260 <YOS>

A;Cross-references: UNIPARC:UPI000006FC05; EMBL:X55543; NID:G287644; PIDN:CAA39149.1; PMID:10617197

R;Ponath, P.D.; Fass, D.; Liou, H.C.; Glimcher, L.H.; Strominger, J.L.

J. Biol. Chem. 268, 17074-17082, 1993

A;Title: The regulatory gene, hXBP-1, and its target, HLA-DRA, utilize both common and distinct enhancers

A;Reference number: A47490; MUID:93352484; PMID:8349596

A;Accession: A47490

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-32, 'GQA', 35-128, 'T', 130-192, 'P', 194-198, 'R', 200-260 <PON>

A;Cross-references: UNIPARC:UPI000017330E; GB:L13850

A;Note: authors translated the codon ACC for residue 130 as Asn, AAG for residue 151 as Lys

C;Genetics:

A;Gene: GDB:XBP1; XBP2

A;Cross-references: GDB:131393; OMIM:194355

A;Map position: 22pter-22qter

A;Introns: 76/3; 108/3; 151/3; 200/3

C;Superfamily: X box-binding protein 1; fos/jun DNA-binding domain homology

C;Keywords: DNA binding; leucine zipper; nucleus; transcription regulation

F;64-104/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 27.0%; Score 87.5; DB 1; Length 260;

Best Local Similarity 46.3%; Pred. No. 0.3;

Matches 19; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

QY 24 RKRAK-TEDEKQRRRIERVLRNRAAAQTSSRERKLEMEKLE 63

DB 58 RKQRLTHLSPEKALRRKLNRAAAQTARDRKARMEKLE 98

RESULT 9

JC7300

tax-responsive element-binding protein 5 - mouse

C;Species: Mus musculus (house mouse)

C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 09-Jul-2004

C;Accession: JC7300

R;Masaki, T.; Noguchi, H.; Kobayashi, M.; Yoshida, M.; Takamatsu, K.

DNA Res. 7, 187-193, 2000

A;Title: Isolation and characterization of the gene encoding mouse tax-responsive element-binding protein 5

A;Reference number: JC7300

A;Accession: JC7300

A;Molecule type: mRNA

A;Residues: 1-266 <MAS>

A;Cross-references: UNIPROT:Q9ESS3; UNIPARC:UPI00000E73B0; DDBJ:AB036745

C;Genetics:

A;Gene: trebs

C;Superfamily: X box-binding protein 1; fos/jun DNA-binding domain homology

C;Keywords: leucine zipper; transcription factor

Query Match 27.0%; Score 87.5; DB 2; Length 266;

Best Local Similarity 46.3%; Pred. No. 0.31;

Matches 19; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

QY 24 RKRAK-TEDEKQRRRIERVLRNRAAAQTSSRERKLEMEKLE 63

DB 51 RKQRLTHLSPEKALRRKLNRAAAQTARDRKARMEKLE 91

RESULT 10

JC4857

hepatocarcinogenesis-related transcription factor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 15-Aug-1996 #sequence_revision 18-Oct-1996 #text_change 09-Jul-2004

C;Accession: JC4857

R;Kishimoto, T.; Kokura, K.; Kumagai, Y.; Makino, Y.; Tamura, T.

Biochem. Biophys. Res. Commun. 224, 746-751, 1996

A;Title: HTF: A b-zip transcription factor that is closely related to the human XBP/TREB

A;Reference number: JC4857

A;Accession: JC4857

A;Molecule type: mRNA

A;Residues: 1-267 <KIS>

A;Cross-references: UNIPROT:Q9R1S4; UNIPARC:UPI00000B80AD

C;Comment: This is a basic-leucine zipper type transcription factor involved in hepatocel

C;Genetics:

A;Gene: htfl

C;Superfamily: X box-binding protein 1; fos/jun DNA-binding domain homology

C;Keywords: leucine zipper; transcription factor

F;58-98/Domain: fos/jun DNA-binding domain homology <FJD>

F;89-126/Region: leucine zipper motif

Query Match 27.0%; Score 87.5; DB 2; Length 267;

Best Local Similarity 46.3%; Pred. No. 0.31;

Matches 19; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

QY 24 RKRAK-TEDEKQRRRIERVLRNRAAAQTSSRERKLEMEKLE 63

DB 52 RKQRLTHLSPEKALRRKLNRAAAQTARDRKARMEKLE 92

RESULT 11

T00759

hypothetical protein At2g40950 [imported] - Arabidopsis thaliana

N;Alternate names: hypothetical protein T20B5.15

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 31-Dec-2004

C;Accession: T00759; H04835

R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, A.

submitted to the EMBL Data Library, November 1997

A;Description: Arabidopsis thaliana chromosome II BAC T20B5 genomic sequence.

A;Reference number: Z14159

A;Accession: T00759

A;Status: translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-600 <RQU>
A:Cross-references: UNIPROT:O22208; UNIPARC:UPI0000179D30; EMBL:AC002409; NID:g2623294;
A:Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Roundley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Mofat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: H84835
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-600 <STO>
A:Cross-references: UNIPARC:UPI0000179D30; GB:AE002093; NID:g2623309; PIDN:AAB86455.1; C
C:Genetics:
A:Gene: T20B5.15; At2g40950
A:Map position: 2
C:Superfamily: bZIP transcription factor; fos/jun DNA-binding domain homology

Query Match 26.5%; Score 86; DB 2; Length 600;
Best Local Similarity 30.8%; Pred. No. 0.91;
Matches 20; Conservative 18; Mismatches 21; Indels 6; Gaps 1;

QY 5 KKRKSGOELPVKTNLPKRKAKTED-----EKEORRIERVLRNRAAAQTSRERKRLLE 58
DB 193 KRKKEIDELTDESRSKYRSGEDADASAVTGEDEKKRRLMRNESAQLSRQRKKHY 252

QY 59 MEKLE 63
DB 253 VEELE 257

RESULT 12
T32750
hypothetical protein F57B10.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession: T32750
R;Greco, T.; Elliott, G.; Keppler, D.
submitted to the EMBL Data Library, December 1997
A:Description: The sequence of C. elegans cosmid F57B10.
A:Reference number: Z21219
A:Accession: T32750
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-688 <GRE>
A:Cross-references: UNIPROT:O44743; UNIPARC:UPI0000079D06; EMBL:AF039713; PIDN:AAB96719.
A:Experimental source: strain Bristol N2; clone F57B10
C:Genetics:
A:Gene: CESP.F57B10.1
A:Map position: 1
A:Introns: 54/1; 150/3; 196/1; 306/3; 383/3; 627/3

Query Match 25.6%; Score 83; DB 2; Length 688;
Best Local Similarity 48.5%; Pred. No. 2;
Matches 16; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

QY 31 DEKEORRIERVLRNRAAAQTSRERKLEMEKLE 63
DB 281 EERDLKIRRKIRNKESAQTSRKRKQDYIEQLE 313

RESULT 13
A42140
box B-binding factor-2 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A42140; S24542
R;Abel, T.; Bhatt, R.; Maniatis, T.
Genes Dev. 6, 466-480, 1992
A>Title: A Drosophila CREB/ATF transcriptional activator binds to both fat body- and liv

A;Reference number: A42140; MUID:92192458; PMID:1532159
A:Accession: A42140
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-515 <ABE>
A:Cross-references: UNIPROT:P29747; UNIPARC:UPI00001283CE; EMBL:X64429; NID:g11063; PIDN:
A:Note: sequence extracted from NCBI backbone (NCBIP:88160)
C:Genetics:
A:Gene: FlyBase:Bbbf2
A:Cross-references: FlyBase:FBgn0004848
C:Superfamily: cAMP-responsive element-binding transcription activator; fos/jun DNA-binding
C:Keywords: DNA binding; nucleus; transcription regulation
F:432-475/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 25.5%; Score 82.5; DB 2; Length 515;
Best Local Similarity 33.9%; Pred. No. 1.7;
Matches 20; Conservative 15; Mismatches 19; Indels 5; Gaps 2;

QY 5 KKRKSGOELPVKTNLPKRKAKTEDEKEORRIERVLRNRAAAQTSRERKLEMEKLE 63
DB 416 EKRTLLAEGYPIQ-KLPLTK---AEKSLKIRKIKNKISAQESRRKKEYMDQLE 469

RESULT 14
A44494
cAMP-responsive element-binding transcription activator CREB-A - fruit fly (Drosophila me
C:Species: Drosophila melanogaster
C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A44494
R;Smolik, S.M.; Rose, R.E.; Goodman, R.H.
Mol. Cell. Biol. 12, 4123-4131, 1992
A>Title: A cyclic AMP-responsive element-binding transcriptional activator in Drosophila
A:Reference number: A44494; MUID:92375081; PMID:1508208
A:Accession: A44494
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-516 <SMO>
A:Cross-references: UNIPROT:P29747; UNIPARC:UPI0000078073; GB:M87038; NID:g157118; PID:g
A:Note: sequence extracted from NCBI backbone (NCBIN:111718, NCBIP:111719)
C:Genetics:
A:Gene: FlyBase:CrebA
A:Cross-references: FlyBase:FBgn0004396
C:Superfamily: cAMP-responsive element-binding transcription activator; fos/jun DNA-binding
F:433-476/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 25.5%; Score 82.5; DB 2; Length 516;
Best Local Similarity 33.9%; Pred. No. 1.7;
Matches 20; Conservative 15; Mismatches 19; Indels 5; Gaps 2;

QY 5 KKRKSGOELPVKTNLPKRKAKTEDEKEORRIERVLRNRAAAQTSRERKLEMEKLE 63
DB 417 EKRTLLAEGYPIQ-KLPLTK---AEKSLKIRKIKNKISAQESRRKKEYMDQLE 470

RESULT 15
T03373
probable G-box binding factor 8 - rice
C:Species: Oryza sativa (rice)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 31-Dec-2004
C:Accession: T03373
R;Nakagawa, H.; Ohmura, K.; Hattori, T.
Plant J. 9, 217-227, 1996
A>Title: A rice bZIP protein, designated OSB28, is rapidly induced by abscisic acid.
A:Reference number: Z14906; MUID:96417817; PMID:8820608
A:Accession: T03373
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: mRNA
A:Residues: 1-360 <NAK>
A:Cross-references: UNIPROT:Q40645; UNIPARC:UPI00000A8257; EMBL:U42208; NID:g1147631; PI
A:Experimental source: cv. Nipponbare
C:Superfamily: BZIP G-box-binding protein; fos/jun DNA-binding domain homology

Query Match 25.0%; Score 81; DB 2; Length 360;

Best Local Similarity 36.1%; Pred. No. 1.7;
 Matches 22; Conservative 10; Mismatches 23; Indels 6; Gaps 1;
 QY 2 KPAKKKSGQELVPKTNLPKRKAKTEDEKESQRIERVLRNRAAAQTSRERKRLMEK 61
 Db 199 KPDVSTASDFRVIATPVTEVP-----TKDDKSKRERRKQSNRESARRSLRKAETEE 252
 QY 62 L 62
 Db 253 L 253

Search completed: November 22, 2005, 15:49:34
 Job time : 29 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: November 22, 2005, 15:31:54 ; Search time 159 Seconds
(without alignments)
176.857 Million cell updates/sec

Title: US-10-663-450-6_COPY_53_116
Perfect score: 324
Sequence: 1 KRPAAKRRKSGWQLPVPKTV.....RAAAQTSRRKRLEMEKLES 64

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*
9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	324	100.0	349	8 ADS12805	Ads12805 Aspergill
2	324	100.0	349	8 ADS12803	Ads12803 Aspergill
3	324	100.0	349	8 ADS12863	Ads12863 Aspergill
4	324	100.0	350	4 AAB82976	Aab82976 Aspergill
5	324	100.0	350	5 AAE15372	Aae15372 Aspergill
6	320	98.8	64	5 AAE15374	Aae15374 Aspergill
7	317	97.8	342	4 AAB82977	Aab82977 Aspergill
8	317	97.8	342	5 AAE15381	Aae15381 Aspergill
9	317	97.8	342	8 ADS12818	Ads12818 Aspergill
10	317	97.8	386	5 AAE15379	Aae15379 Aspergill
11	317	97.8	386	8 ADS12815	Ads12815 Aspergill
12	283	87.3	64	5 AAE15373	Aae15373 Trichoder
13	283	87.3	450	4 AAB82975	Aab82975 Trichoder
14	283	87.3	451	5 AAE15371	Aae15371 Trichoder
15	283	87.3	451	8 ADS12801	Ads12801 Trichoder
16	283	87.3	451	8 ADS12804	Ads12804 Trichoder
17	277	85.5	409	7 ABO43144	Abo43144 A. thalia
18	277	85.5	409	7 ADB31925	Adb31925 Plant (A.
19	277	85.5	409	8 ADO02271	Ado02271 Thalecres
20	171.5	52.9	84	5 ABP02534	Abp02534 Human ORF
21	171.5	52.9	200	8 ADS43437	Ads43437 Bacterial
22	171.5	52.9	230	2 AAW53806	Aaw53806 Transcrip
23	171.5	52.9	230	8 ADT87049	Adt87049 Yeast Str
24	171.5	52.9	230	2 AAW53807	Aaw53807 Transcrip

25	170	52.5	68	5	AAE15382	Aae15382 Yeast HAC
26	170	52.5	68	8	ADS12859	Ads12859 Saccharom
27	116	35.8	168	5	AU93013	Aau93013 Arabidops
28	116	35.8	168	7	ADD30174	Add30174 Plant yie
29	116	35.8	168	8	ADI43893	Adi43893 Plant tra
30	116	35.8	211	3	AAG08861	Aag08861 Arabidops
31	111.5	34.4	192	4	AAB82614	Aab82614 Maize roo
32	111.5	34.4	192	4	AAB82615	Aab82615 Maize roo
33	111.5	34.4	192	4	AAB82616	Aab82616 Maize roo
34	111.5	34.4	192	4	AAG66525	Aag66525 Maize roo
35	111.5	34.4	192	4	AAG66526	Aag66526 Maize roo
36	109.5	33.8	185	9	ADM17162	Adm17162 Eucalyptu
37	106.5	32.9	170	8	ADM48147	Adm48147 Polypepti
38	106	32.7	163	9	ADM17580	Adm17580 Pinus rad
39	100	30.9	143	3	AAB33151	Aab33151 Pinus rad
40	98	30.2	120	3	AAG27808	Aag27808 Arabidops
41	98	30.2	135	3	AAG07181	Aag07181 Arabidops
42	98	30.2	149	3	AAG07180	Aag07180 Arabidops
43	98	30.2	149	3	AAG27807	Aag27807 Arabidops
44	98	30.2	188	3	AAG27806	Aag27806 Arabidops
45	93.5	28.9	672	8	ADX95805	Adx95805 Plant ful

ALIGNMENTS

RESULT 1
ADS12805
ID ADS12805 standard; protein; 349 AA.

XX ADS12805;
AC
XX
DT 16-DEC-2004 (first entry)
XX
DE Aspergillus nidulans hacA DNA binding domain.
XX
KW unfolded protein response; UPR; HAC1; PTC2;
KW unfolded protein response modulator; enzyme production; hacA; chaperone;
KW foldase; DNA binding domain.
XX

OS Emericella nidulans.

PN US2004186070-A1.

XX 23-SEP-2004.

XX 15-SEP-2003; 2003US-00663450.

XX 24-MAR-2000; 2000US-00534692.

PR 23-MAR-2001; 2001US-00816277.

(GEMV) GENENCOR INT INC.

Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;

WPI; 2004-707924/69.

Increasing the secretion of a heterologous protein, such as a therapeutic or an industrial enzyme, in genetically modified eukaryotic cells by inducing an elevated unfolded protein response (UPR).

Example 3; SEQ ID NO 6; 83pp; English.

The invention describes a method of increasing the secretion of a heterologous protein in a eukaryotic cell, comprising inducing an elevated unfolded protein response (UPR). Also described are: an isolated nucleic acid encoding a HAC1 protein, where the HAC1 protein induces unfolded protein response and has less than 50% similarity to yeast HAC1 protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1 protein induces unfolded protein response and where the HAC1 protein comprises a DNA binding region that has greater than 70% similarity to the DNA binding region of filamentous fungi HAC1 protein; a protein having unfolded protein response inducing activity and having greater

CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
CC respectively) as given in the specification; a protein having an amino
CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
CC response and has at least 70% similarity to a fully defined amino acid
CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
CC specification; an isolated nucleic acid encoding a IRE1 protein that
CC modulates unfolded protein response and has at least 60% similarity to a
CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
CC having unfolded protein response modulating activity and having greater
CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
CC heterologous nucleic acid encoding a protein having unfolded protein
CC response modulating activity and a heterologous nucleic acid encoding a
CC protein of interest to be secreted. The methods and compositions of
CC genetically manipulating cells to have an elevated unfolded protein
CC response (UPR) resulting in an increased capacity to produce secreted
CC proteins, are useful in e.g. production of therapeutic or industrial
CC enzymes. This is the amino acid sequence of *Aspergillus nidulans* HAC1
CC chaperone and foldase DNA binding domain.
XX
SQ Sequence 349 AA;

Query Match 100.0%; Score 324; DB 8; Length 349;
Best Local Similarity 100.0%; Pred. No. 9.9e-30;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKPAAKRRKSGOELPVPKTNLPKRKTEDEKEQRIERVLRNRAAAQTSRKRLEME 60
|||||
Db 53 KKPAAKRRKSGOELPVPKTNLPKRKTEDEKEQRIERVLRNRAAAQTSRKRLEME 112

QY 61 KLES 64
|||||
Db 113 KLES 116

RESULT 2
ADSI12803
ID ADSI12803 standard; protein; 349 AA.
AC ADSI12803;
DT 16-DEC-2004 (first entry)
DE *Aspergillus nidulans* hacA chaperone and foldase #1.
KW unfolded protein response; UPR; HAC1; PTC2;
KW unfolded protein response modulator; enzyme production; hacA; chaperone;
KW foldase; enzyme.
XX
OS *Emericella nidulans*.
XX
PN US2004186070-A1.
XX
XX
XX 23-SEP-2004.
XX
XX 15-SEP-2003; 2003US-00663450.
XX
XX 24-MAR-2000; 2000US-00534692.
XX 23-MAR-2001; 2001US-00816277.
XX
XX (GEMV) GENENCOR INT INC.
XX
XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX WPI; 2004-707924/69.
XX N-PSDB; ADSI12802.
XX
XX Increasing the secretion of a heterologous protein, such as a therapeutic
XX or an industrial enzyme, in genetically modified eukaryotic cells by
XX inducing an elevated unfolded protein response (UPR).
XX
XX Example 3; SEQ ID NO 4; 83pp; English.

XX The invention describes a method of increasing the secretion of a
CC heterologous protein in a eukaryotic cell, comprising inducing an
CC elevated unfolded protein response (UPR). Also described are: an isolated
CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
CC unfolded protein response and has less than 50% similarity to yeast HAC1
CC protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
CC protein induces unfolded protein response and where the HAC1 protein
CC comprises a DNA binding region that has greater than 70% similarity to
CC the DNA binding region of filamentous fungi HAC1 protein; a protein
CC having unfolded protein response inducing activity and having greater
CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
CC respectively) as given in the specification; a protein having an amino
CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
CC response and has at least 70% similarity to a fully defined amino acid
CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
CC specification; an isolated nucleic acid encoding a IRE1 protein that
CC modulates unfolded protein response and has at least 60% similarity to a
CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
CC having unfolded protein response modulating activity and having greater
CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
CC heterologous nucleic acid encoding a protein having unfolded protein
CC response modulating activity and a heterologous nucleic acid encoding a
CC protein of interest to be secreted. The methods and compositions of
CC genetically manipulating cells to have an elevated unfolded protein
CC response (UPR) resulting in an increased capacity to produce secreted
CC proteins, are useful in e.g. production of therapeutic or industrial
CC enzymes. This is the amino acid sequence of *Aspergillus nidulans* HAC1
CC chaperone and foldase. Note: This sequence differs from ADSI12863 given in
CC figure 8.
XX
SQ Sequence 349 AA;

Query Match 100.0%; Score 324; DB 8; Length 349;
Best Local Similarity 100.0%; Pred. No. 9.9e-30;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKPAAKRRKSGOELPVPKTNLPKRKTEDEKEQRIERVLRNRAAAQTSRKRLEME 60
|||||
Db 53 KKPAAKRRKSGOELPVPKTNLPKRKTEDEKEQRIERVLRNRAAAQTSRKRLEME 112

QY 61 KLES 64
|||||
Db 113 KLES 116

RESULT 3
ADSI12863
ID ADSI12863 standard; protein; 349 AA.
XX
AC ADSI12863;
XX
XX 16-DEC-2004 (first entry)
XX
XX *Aspergillus nidulans* hacA chaperone and foldase #2.
DE
XX
XX unfolded protein response; UPR; HAC1; PTC2;
XX unfolded protein response modulator; enzyme production; HAC1; chaperone;
KW foldase; enzyme.
XX
XX *Emericella nidulans*.
OS
XX
XX US2004186070-A1.
PN
XX
XX 23-SEP-2004.
XX
XX 15-SEP-2003; 2003US-00663450.
XX
XX 24-MAR-2000; 2000US-00534692.
XX 23-MAR-2001; 2001US-00816277.
XX
XX (GEMV) GENENCOR INT INC.

PA (GEMV) GENENCOR INT INC.
 XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
 XX WPI; 2004-707924/69.
 XX N-PSDB; ADS12802.
 XX
 PT Increasing the secretion of a heterologous protein, such as a therapeutic
 PT or an industrial enzyme, in genetically modified eukaryotic cells by
 PT inducing an elevated unfolded protein response (UPR).
 XX
 XX Example 3; Fig 8; 83pp; English.
 XX
 CC The invention describes a method of increasing the secretion of a
 CC heterologous protein in a eukaryotic cell, comprising inducing an
 CC elevated unfolded protein response (UPR). Also described are: an isolated
 CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
 CC unfolded protein response and has less than 50% similarity to yeast HAC1
 CC protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
 CC protein induces unfolded protein response and where the HAC1 protein
 CC comprises a DNA binding region that has greater than 70% similarity to
 CC the DNA binding region of filamentous fungi HAC1 protein; a protein
 CC having unfolded protein response inducing activity and having greater
 CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
 CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
 CC respectively) as given in the specification; a protein having an amino
 CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
 CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
 CC response and has at least 70% similarity to a fully defined amino acid
 CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
 CC specification; an isolated nucleic acid encoding a IRE1 protein that
 CC modulates unfolded protein response and has at least 60% similarity to a
 CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
 CC having unfolded protein response modulating activity and having greater
 CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
 CC heterologous nucleic acid encoding a protein having unfolded protein
 CC response modulating activity and a heterologous nucleic acid encoding a
 CC protein of interest to be secreted. The methods and compositions of
 CC genetically manipulating cells to have an elevated unfolded protein
 CC response (UPR) resulting in an increased capacity to produce secreted
 CC proteins, are useful in e.g. production of therapeutic or industrial
 CC enzymes. This is the amino acid sequence of Aspergillus nidulans HAC1
 CC chaparrone and foldase. Note: This sequence differs from ADS12803 given in
 CC the sequence listing.
 XX
 SQ Sequence 349 AA;
 Query Match 100.0%; Score 324; DB 8; Length 349;
 Best Local Similarity 100.0%; Pred. No. 9.9e-30;
 Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KKPAAKKKSGQELPVPKTNLPKPKRAKTEDEKEQRRIERVLNRNRAAAQTSRERKLEME 60
 DB 53 KKPAAKKKSGQELPVPKTNLPKPKRAKTEDEKEQRRIERVLNRNRAAAQTSRERKLEME 112
 QY 61 KLES 64
 DB 113 KLES 116
 RESULT 4
 AAB82976
 ID AAB82976 standard; protein; 350 AA.
 XX
 AC AAB82976;
 XX
 XX 11-SEP-2003 (revised)
 DT 21-DEC-2001 (first entry)
 XX
 DE Aspergillus nidulans hacA, involved in unfolded protein response.
 XX
 KW HacA; transcription factor; unfolded protein response; protein secretion.
 XX

OS Emericella nidulans.
 XX Key Location/Qualifiers
 FH Domain 53..116
 FT /label= DNA binding domain
 XX
 XX WO200172783-A2.
 XX 04-OCT-2001.
 XX 23-MAR-2001; 2001WO-US009401.
 XX 24-MAR-2000; 2000US-00534692.
 XX (GEMV) GENENCOR INT INC.
 XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
 XX WPI; 2001-626252/72.
 XX N-PSDB; AAH26932.
 XX
 PT Increasing secretion of heterologous proteins e.g. lipase and cellulase
 PT in eukaryotic cells useful in industry to increase production and
 PT facilitate purification, by inducing an elevated unfolded protein
 PT response.
 XX
 XX Claim 54; Fig 8A-B; 89pp; English.
 XX
 CC The present sequence is that of the hacA protein of Aspergillus nidulans,
 CC as deduced from the newly isolated hacA gene (see AAH26932). HacA protein
 CC is a transcription factor involved in the unfolded protein response
 CC (UPR). The invention provides methods for increasing the secretion of a
 CC heterologous protein in a cell by inducing an elevated UPR. This can be
 CC achieved by modulating the activity of HAC1 (or hacA), PTC2 or IRE1 in
 CC the cell, e.g. by gene overexpression. The cell from which the protein is
 CC secreted can be any cell having an UPR, such as mammalian cells, insect
 CC cells, yeast and filamentous fungi. The protein of interest can be any
 CC secreted protein such as a therapeutic protein or an industrial enzyme,
 CC e.g. lipase, cellulase, endoglucanase-H, protease, carbonhydratase,
 CC reductase, oxidase, isomerase, transferase, kinase, phosphatase, alpha-
 CC amylase, glucosylase, lignocellulose hemicellulase, pectinase and
 CC ligninase (claimed). (Updated on 11-SEP-2003 to standardise OS field)
 XX
 SQ Sequence 350 AA;
 Query Match 100.0%; Score 324; DB 4; Length 350;
 Best Local Similarity 100.0%; Pred. No. 9.9e-30;
 Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KKPAAKKKSGQELPVPKTNLPKPKRAKTEDEKEQRRIERVLNRNRAAAQTSRERKLEME 60
 DB 54 KKPAAKKKSGQELPVPKTNLPKPKRAKTEDEKEQRRIERVLNRNRAAAQTSRERKLEME 113
 QY 61 KLES 64
 DB 114 KLES 117
 RESULT 5
 AAEL15372
 ID AAEL15372 standard; protein; 350 AA.
 XX
 AC AAEL15372;
 XX
 XX 29-AUG-2003 (revised)
 DT 07-MAR-2002 (first entry)
 XX
 DE Aspergillus nidulans hacA protein.
 XX
 KW Heterologous protein secretion; unfolded protein response; UPR; lipase;
 KW cellulase; carbohydrase; industry; purification; hacA protein.
 XX
 XX Emericella nidulans.
 OS

```

XX Key Location/Qualifiers
FH Binding-site 53..116
FT /label= DNA-binding_domain
FT
XX
XX US2001034045-A1.
XX
XX 25-OCT-2001.
XX
XX 23-MAR-2001; 2001US-00816277.
XX
XX 24-MAR-2000; 2000US-00534692.
XX
XX (GEMV ) GENENCOR INT INC.
XX
XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX WPI; 2002-033728/04.
XX
XX Increasing secretion of heterologous proteins e.g. lipase and cellulase
XX in eukaryotic cells useful in industry to increase production and
XX facilitate purification, by inducing an elevated unfolded protein
XX response.
XX
XX Claim 39; Fig 8; 56pp; English.
XX
XX The present invention relates to methods for increasing the secretion of
XX heterologous protein in eukaryotic cells by inducing an elevated unfolded
XX protein response (UPR). The method involves inducing the elevated UPR by
XX increasing the presence of proteins such as HAC1, HACA, PTC2 or LRE1 in
XX cells. The method and sequences are useful for increasing the secretion
XX of heterologous proteins (e.g. lipase, cellulase, carboxylase) in
XX eukaryotic cells useful in industry to increase protein yields and to
XX facilitate purification. The present sequence is Aspergillus nidulans
XX hacA protein. (Updated on 29-AUG-2003 to standardise OS field)
XX
XX Sequence 350 AA;
XX
XX Query Match 100.0%; Score 324; DB 5; Length 350;
XX Best Local Similarity 100.0%; Pred. No. 9.9e-30;
XX Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 KXPAKKRKSQGQELPVPKTNLPKRAKTEDEKEQRRIERVLNRNAAAQTSRKRKLEME 60
XX Db 54 KXPAKKRKSQGQELPVPKTNLPKRAKTEDEKEQRRIERVLNRNAAAQTSRKRKLEME 113
XX
XX QY 61 KLES 64
XX Db 114 KLES 117
XX
XX RESULT 6
XX AA015374
XX ID AA015374 standard; protein; 64 AA.
XX
XX AC AA015374;
XX
XX DT 29-AUG-2003 (revised)
XX DT 07-MAR-2002 (first entry)
XX
XX DE Aspergillus nidulans hacA protein DNA binding domain.
XX
XX KW Heterologous protein secretion; unfolded protein response; UPR; lipase;
XX KW cellulase; carboxylase; industry; purification; DNA binding domain;
XX KW hacA protein.
XX
XX OS Emericella nidulans.
XX
XX PN US2001034045-A1.
XX
XX PD 25-OCT-2001.
XX

```

```

PF 23-MAR-2001; 2001US-00816277.
XX
XX 24-MAR-2000; 2000US-00534692.
XX
XX (GEMV ) GENENCOR INT INC.
XX
XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX WPI; 2002-033728/04.
XX
XX Increasing secretion of heterologous proteins e.g. lipase and cellulase
XX in eukaryotic cells useful in industry to increase production and
XX facilitate purification, by inducing an elevated unfolded protein
XX response.
XX
XX Example 3; Fig 10; 56pp; English.
XX
XX The present invention relates to methods for increasing the secretion of
XX heterologous protein in eukaryotic cells by inducing an elevated UPR by
XX protein response (UPR). The method involves inducing the elevated UPR by
XX increasing the presence of proteins such as HAC1, HACA, PTC2 or LRE1 in
XX cells. The method and sequences are useful for increasing the secretion
XX of heterologous proteins (e.g. lipase, cellulase, carboxylase) in
XX eukaryotic cells useful in industry to increase protein yields and to
XX facilitate purification. The present sequence is Aspergillus nidulans
XX hacA protein DNA binding domain. (Updated on 29-AUG-2003 to standardise
XX OS field)
XX
XX Sequence 64 AA;
XX
XX Query Match 98.8%; Score 320; DB 5; Length 64;
XX Best Local Similarity 100.0%; Pred. No. 4.6e-30;
XX Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 KXPAKKRKSQGQELPVPKTNLPKRAKTEDEKEQRRIERVLNRNAAAQTSRKRKLEME 60
XX Db 2 KXPAKKRKSQGQELPVPKTNLPKRAKTEDEKEQRRIERVLNRNAAAQTSRKRKLEME 61
XX
XX QY 61 KLE 63
XX Db 62 KLE 64
XX
XX RESULT 7
XX AA082977
XX ID AA082977 standard; protein; 342 AA.
XX
XX AC AA082977;
XX
XX DT 11-SEP-2003 (revised)
XX DT 21-DEC-2001 (first entry)
XX
XX DE Aspergillus niger hacA, involved in unfolded protein response.
XX
XX KW HAC; transcription factor; unfolded protein response; protein secretion.
XX
XX OS Aspergillus awamori.
XX
XX FH Key Location/Qualifiers
XX FT Domain 45..109
XX FT /label= DNA binding domain
XX
XX PN WO200172783-A2.
XX
XX PD 04-OCT-2001.
XX
XX PF 23-MAR-2001; 2001WO-US009401.
XX
XX PR 24-MAR-2000; 2000US-00534692.
XX
XX (GEMV ) GENENCOR INT INC.
XX
XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX

```

XX WPI; 2001-626252/72.
 DR N-PSDB; AAH26933.
 XX
 PT Increasing secretion of heterologous proteins e.g. lipase and cellulase
 PT in eukaryotic cells useful in industry to increase production and
 PT facilitate purification, by inducing an elevated unfolded protein
 PT response.
 XX
 PS Claim 54; Fig 28A-C; 89pp; English.
 XX
 CC The present sequence is that of the haca protein of *Aspergillus niger*
 CC var. awamori, as deduced from haca cDNA (see AAH26933). Haca protein is a
 CC transcription factor involved in the unfolded protein response (UPR).
 CC Overexpression of an inducing form of haca enables production of higher
 CC levels of secreted heterologous proteins in *A. niger*. The invention
 CC provides methods for increasing the secretion of a heterologous protein
 CC in a cell by inducing an elevated UPR. This can be achieved by modulating
 CC the activity of HAC1 (or haca), PTC2 or IRE1 in the cell, e.g. by gene
 CC overexpression. The cell from which the protein is secreted can be any
 CC cell having an UPR, such as mammalian cells, insect cells, yeast and
 CC filamentous fungi. The protein of interest can be any secreted protein
 CC such as a therapeutic protein or an industrial enzyme, e.g. lipase,
 CC cellulase, endoglucanase-H, protease, carboxylatase, reductase, oxidase,
 CC isomerase, transferase, kinase, phosphatase, alpha-amylase, glucoamylase,
 CC lignocellulose hemicellulase, pectinase and ligninase (claimed). (Updated
 CC on 11-SEP-2003 to standardise OS field)
 XX
 SQ Sequence 342 AA;
 Query Match 97.8%; Score 317; DB 4; Length 342;
 Best Local Similarity 96.9%; Pred. No. 6.5e-29;
 Matches 62; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 KKPAAKRSWGQELPVPKTNLPKRAKTEDEKEQRIERVLRNRAAAQTSRKRLEME 60
 DB 47 KKPVKRKSGQELPVPKTNLPKRAKTEDEKEQRIERVLRNRAAAQTSRKRLEME 106
 QY 61 KLES 64
 DB 107 KLEN 110
 XX
 RESULT 8
 ID AAEE15381 standard; protein; 342 AA.
 AC AAEE15381;
 DT 07-MAR-2002 (first entry)
 XX
 DE *Aspergillus niger* var. awamori haca protein #3.
 XX
 KW Heterologous protein secretion; unfolded protein response; UPR; lipase;
 KW cellulase; carboxylatase; industry; purification; haca protein.
 XX
 OS *Aspergillus niger*.
 XX
 PN US2001034045-A1.
 PD 25-OCT-2001.
 XX
 PF 23-MAR-2001; 2001US-00816277.
 XX
 PR 24-MAR-2000; 2000US-00534692.
 XX
 PA (GENV) GENENCOR INT INC.
 XX
 PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
 XX
 DR WPI; 2002-033728/04.
 XX
 PT Increasing secretion of heterologous proteins e.g. lipase and cellulase

PT in eukaryotic cells useful in industry to increase production and
 PT facilitate purification, by inducing an elevated unfolded protein
 PT response.
 XX
 PS Claim 39; Fig 28; 56pp; English.
 XX
 CC The present invention relates to methods for increasing the secretion of
 CC heterologous protein in eukaryotic cells by inducing an elevated unfolded
 CC protein response (UPR). The method involves inducing the elevated UPR by
 CC increasing the presence of proteins such as HAC1, HACA, PTC2 or IRE1 in
 CC cells. The method and sequences are useful for increasing the secretion
 CC of heterologous proteins (e.g. lipase, cellulase, carboxylatase) in
 CC eukaryotic cells useful in industry to increase protein yields and to
 CC facilitate purification. The present sequence is *Aspergillus niger* var.
 CC awamori haca protein
 XX
 SQ Sequence 342 AA;
 Query Match 97.8%; Score 317; DB 5; Length 342;
 Best Local Similarity 96.9%; Pred. No. 6.5e-29;
 Matches 62; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 KKPAAKRSWGQELPVPKTNLPKRAKTEDEKEQRIERVLRNRAAAQTSRKRLEME 60
 DB 47 KKPVKRKSGQELPVPKTNLPKRAKTEDEKEQRIERVLRNRAAAQTSRKRLEME 106
 QY 61 KLES 64
 DB 107 KLEN 110
 XX
 RESULT 9
 ID ADS12818 standard; protein; 342 AA.
 AC ADS12818;
 XX
 DT 16-DEC-2004 (first entry)
 XX
 DE *Aspergillus nidulans* haca polypeptide seqid 19.
 XX
 KW unfolded protein response; UPR; HAC1; PTC2;
 KW unfolded protein response modulator; enzyme production; haca; chaperone;
 KW foldase.
 XX
 OS *Emmericella nidulans*.
 XX
 PN US2004186070-A1.
 XX
 PD 23-SEP-2004.
 XX
 PF 15-SEP-2003; 2003US-00663450.
 XX
 PR 24-MAR-2000; 2000US-00534692.
 PR 23-MAR-2001; 2001US-00816277.
 XX
 PA (GENV) GENENCOR INT INC.
 XX
 PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
 XX
 DR WPI; 2004-707924/69.
 XX
 PT Increasing the secretion of a heterologous protein, such as a therapeutic
 PT or an industrial enzyme, in genetically modified eukaryotic cells by
 PT inducing an elevated unfolded protein response (UPR).
 XX
 PS Example 12; SEQ ID NO 19; 83pp; English.
 XX
 CC The invention describes a method of increasing the secretion of a
 CC heterologous protein in a eukaryotic cell, comprising inducing an
 CC elevated unfolded protein response (UPR). Also described are: an isolated
 CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
 CC unfolded protein response and has less than 50% similarity to yeast HAC1

CC protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
 CC protein induces unfolded protein response and where the HAC1 protein to
 CC comprises a DNA binding region that has greater than 70% similarity to
 CC the DNA binding region of filamentous fungi HAC1 protein; a protein
 CC having unfolded protein response inducing activity and having greater
 CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
 CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
 CC respectively) as given in the specification; a protein having an amino
 CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
 CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
 CC response and has at least 70% similarity to a fully defined amino acid
 CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
 CC specification; an isolated nucleic acid encoding a IRE1 protein that
 CC modulates unfolded protein response and has at least 60% similarity to a
 CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
 CC having unfolded protein response modulating activity and having greater
 CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
 CC heterologous nucleic acid encoding a protein having unfolded protein
 CC response modulating activity and a heterologous nucleic acid encoding a
 CC protein of interest to be secreted. The methods and compositions of
 CC genetically manipulating cells to have an elevated unfolded protein
 CC response (UPR) resulting in an increased capacity to produce secreted
 CC proteins, are useful in e.g. production of therapeutic or industrial
 CC enzymes. This is the amino acid sequence of *Aspergillus nidulans* HAC1
 CC chaparrone and foldase.

XX
 SQ Sequence 386 AA;

Query Match 97.8%; Score 317; DB 8; Length 386;
 Best Local Similarity 96.9%; Pred. No. 7.4e-29;
 Matches 62; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KKPAAKRSWGQELPVPKTNLPKRAKTEDEKEQRRIERVLNRAAAQTSSRRKKLEME 60
 DB 91 KKPVKRRKSWGQELPVPKTNLPKRAKTEDEKEQRRIERVLNRAAAQTSSRRKKLEME 150
 OY 61 KLES 64
 DB 151 KLEN 154

RESULT 12
 AAEE15373
 ID AAEE15373 standard; protein; 64 AA.

XX AC AAEE15373;
 XX DT 29-AUG-2003 (revised)
 XX DT 07-MAR-2002 (first entry)
 XX DE Trichoderma reesei HAC1 protein DNA binding domain.

XX KW Heterologous protein secretion; unfolded protein response; UPR; lipase;
 KW cellulase; carboxylase; industry; purification; DNA binding domain;
 KW HAC1 protein.

XX OS Hypocrea jecorina.
 XX PN US2001034045-A1.

XX PD 25-OCT-2001.
 XX PF 23-MAR-2001; 2001US-00816277.
 XX PR 24-MAR-2000; 2000US-00534692.

XX PA (GENV) GENENCOR INT INC.

XX PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MUA;
 XX WPI; 2002-033728/04.

PT Increasing secretion of heterologous proteins e.g. lipase and cellulase

PT in eukaryotic cells useful in industry to increase production and
 PT facilitate purification, by inducing an elevated unfolded protein
 PT response.

XX Example 3; Fig 10; 56pp; English.

XX CC The present invention relates to methods for increasing the secretion of
 CC heterologous protein in eukaryotic cells by inducing an elevated unfolded
 CC protein response (UPR). The method involves inducing the elevated UPR by
 CC increasing the presence of proteins such as HAC1, HACA, PTC2 or IRE1 in
 CC cells. The method and sequences are useful for increasing the secretion
 CC of heterologous proteins (e.g. lipase, cellulase, carboxylase) in
 CC eukaryotic cells useful in industry to increase protein yields and to
 CC facilitate purification. The present sequence is *Trichoderma reesei* HAC1
 CC protein DNA binding domain. (Updated on 29-AUG-2003 to standardise OS
 CC field)

XX SQ Sequence 64 AA;

Query Match 87.3%; Score 283; DB 5; Length 64;
 Best Local Similarity 87.3%; Pred. No. 1.1e-25;
 Matches 55; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 KKPAAKRSWGQELPVPKTNLPKRAKTEDEKEQRRIERVLNRAAAQTSSRRKKLEME 60
 DB 2 KKPVKRRKSWGQELPVPKTNLPKRAKTEDEKEQRRIERVLNRAAAQTSSRRKKLEME 61

OY 61 KLE 63
 DB 62 ALE 64

RESULT 13
 AAB82975
 ID AAB82975 standard; protein; 450 AA.

XX AC AAB82975;
 XX DT 11-SEP-2003 (revised)
 XX DT 21-DEC-2001 (first entry)

XX DE Trichoderma reesei HAC1, involved in unfolded protein response.

XX KW HAC1; transcription factor; unfolded protein response; protein secretion.

XX OS Hypocrea jecorina.

XX FH Key Location/Qualifiers
 XX FT Domain 84..147
 XX FT /label= DNA binding domain

XX PN WO200172783-A2.

XX PD 04-OCT-2001.

XX PF 23-MAR-2001; 2001WO-US009401.

XX PR 24-MAR-2000; 2000US-00534692.

XX PA (GENV) GENENCOR INT INC.

XX PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MUA;

XX WPI; 2001-626252/72.
 XX N-PSDB; AAB26931.

XX PT Increasing secretion of heterologous proteins e.g. lipase and cellulase
 PT in eukaryotic cells useful in industry to increase production and
 PT facilitate purification, by inducing an elevated unfolded protein
 PT response.

XX PS Claim 54; Fig 7A-B; 89pp; English.

CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
CC respectively) as given in the specification; a protein having an amino
CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
CC response and has at least 70% similarity to a fully defined amino acid
CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
CC specification; an isolated nucleic acid encoding a IRE1 protein that
CC modulates unfolded protein response and has at least 60% similarity to a
CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
CC having unfolded protein response modulating activity and having greater
CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
CC heterologous nucleic acid encoding a protein having unfolded protein
CC response modulating activity and a heterologous nucleic acid encoding a
CC protein of interest to be secreted. The methods and compositions of
CC genetically manipulating cells to have an elevated unfolded protein
CC response (UPR) resulting in an increased capacity to produce secreted
CC proteins, are useful in e.g. production of therapeutic or industrial
CC enzymes. This is the amino acid sequence of Trichoderma reesei HAC1
CC chaperone and foldase.
XX
SQ Sequence 451 AA;

Query Match 87.3%; Score 283; DB 8; Length 451;
Best Local Similarity 87.3%; Pred. No. 9.1e-25;
Matches 55; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 KKPAAKRRKSWGQELPVPKTNLPKRAKTEDEKEQRRIRVLRNRAAAQTSSRRKRLEME 60
DB 85 KKPVKRRKSWGQVLPKTNLPKRAKTEDEKEQRRIRVLRNRAAQSSRRKRLEVE 144
QY 61 KLE 63
DB 145 ALR 147

Search completed: November 22, 2005, 15:43:40
Job time : 169 secs

THIS PAGE BLANK (USPTO)

Result No.	Score	Query Match	Length	DB	ID	Description
1	307	95.0	409	2	US-09-533-029-104	Sequence 104, Appl
2	174	53.9	248	2	US-09-248-796A-18840	Sequence 18840, A
3	99	30.7	143	2	US-09-640-211A-1930	Sequence 1930, Ap
4	89	27.6	23	2	US-08-531-642-72	Sequence 72, Appl
5	89	27.6	371	2	US-09-148-545-259	Sequence 259, App
6	89	27.6	371	2	US-09-621-011-259	Sequence 259, App
7	86	26.6	102	2	US-09-640-211A-1046	Sequence 1046, Ap
8	82.5	25.5	611	2	US-09-216-393B-81	Sequence 81, Appl
9	81	25.1	212	2	US-09-489-039A-13074	Sequence 13074, A
10	80.5	24.9	395	2	US-09-247-155-113	Sequence 113, App
11	80.5	24.9	395	2	US-09-513-599C-14	Sequence 14, Appl
12	80.5	24.9	395	2	US-09-471-276-14	Sequence 14, Appl
13	80.5	24.9	395	2	US-09-503-190-113	Sequence 113, App
14	80	24.8	63	2	US-08-107-433-3674	Sequence 3674, Ap
15	78	24.1	551	2	US-08-796-899-29	Sequence 29, Appl
16	77.5	24.0	501	2	US-09-949-016-11281	Sequence 11281, A
17	77	23.8	284	2	US-08-949-016-10789	Sequence 10789, A
18	76.5	23.7	351	2	US-08-870-089B-2	Sequence 2, Appli
19	76.5	23.7	362	2	US-09-949-016-10923	Sequence 10923, A
20	75.5	23.4	463	2	US-09-345-236B-2	Sequence 2, Appli
21	75	23.2	170	2	US-09-248-796A-27838	Sequence 27838, A
22	73.5	22.8	260	2	US-09-538-092-950	Sequence 950, App
23	73.5	22.8	338	1	US-08-218-686-2	Sequence 2, Appli
24	73.5	22.8	338	2	US-08-460-242-2	Sequence 2, Appli
25	73.5	22.8	2058	2	US-09-949-016-6835	Sequence 6835, Ap
26	73.5	22.8	2111	2	US-09-949-016-10199	Sequence 10199, A
27	73	22.6	325	2	US-09-567-031-14	Sequence 14, Appl

```
/ Patent No. 6747137
/ GENERAL INFORMATION:
/ APPLICANT: Keith Weinstein et al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
/ TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 107196.132
/ CURRENT APPLICATION NUMBER: US/09/248,796A
/ CURRENT FILING DATE: 1999-02-12
/ PRIOR FILING DATE: 1999-02-12
/ PRIOR FILING DATE: 1998-02-13
/ PRIOR APPLICATION NUMBER: US 60/074,725
/ PRIOR FILING DATE: 1998-02-13
/ PRIOR APPLICATION NUMBER: US 60/096,409
/ NUMBER OF SEQ ID NOS: 28208
/ SEQ ID NO 18840
/ LENGTH: 248
/ TYPE: PRT
/ ORGANISM: Candida albicans
/ FEATURE:
/ NAME/KEY: UNSURE
/ LOCATION: (221)
/ OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown
US-09-248-796A-18840

Query Match      53.9%; Score 174; DB 2; Length 248;
Best Local Similarity 71.7%; Pred. No. 2.6e-12;
Matches 33; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

QY 19 KTNLPKRAKTEDEKEQRRVERVLNRRRAAQSRRERKEVEALE 64
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 72 KSTLPKRAKTEDEKEQRRVERVLNRRRAAQSRRERKEVEALE 117

RESULT 3
US-09-640-211A-1930
/ Sequence 1930, Application US/09640211A
/ Patent No. 6833446
/ GENERAL INFORMATION:
/ APPLICANT: Wood, Marion
/ APPLICANT: Shenk, Michael A.
/ APPLICANT: McGrath, Annette
/ APPLICANT: Glenn, Matthew
/ TITLE OF INVENTION: Compositions and Methods for the
/ TITLE OF INVENTION: Modification of Gene Transcription
/ FILE REFERENCE: 11000.1021C1U
/ CURRENT APPLICATION NUMBER: US/09/640,211A
/ CURRENT FILING DATE: 2000-08-16
/ NUMBER OF SEQ ID NOS: 2368
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 1930
/ LENGTH: 143
/ TYPE: PRT
/ ORGANISM: Pinus radiata
US-09-640-211A-1930

Query Match      30.7%; Score 99; DB 2; Length 143;
Best Local Similarity 47.6%; Pred. No. 0.00056;
Matches 20; Conservative 9; Mismatches 13; Indels 0; Gaps 0;

QY 23 PPKRAKTEDEKEQRRVERVLNRRRAAQSRRERKEVEALE 64
   |||||: ||||| ||||| ||||| |||||: |||||: |||||:
Db 35 PPKRAADLNAEQRRREARHNRIAAQNSRDKRAQTYME 76

RESULT 4
US-09-831-642-72
/ Sequence 72, Application US/09831642
/ Patent No. 6635751
/ GENERAL INFORMATION:
/ APPLICANT: HAZE, Kyosuke et al.
/ TITLE OF INVENTION: ENDOPLASMIC RETICULUM STRESS TRANSCRIPTION FACTORS ATF6 AND CREB-
/ FILE REFERENCE: 1422-0474P
/ CURRENT APPLICATION NUMBER: US/09/831,642
/ CURRENT FILING DATE: 2001-05-11
```

```
/ NUMBER OF SEQ ID NOS: 83
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 72
/ LENGTH: 23
/ TYPE: PRT
/ ORGANISM: Yeast HacIp
US-09-831-642-72

Query Match      27.6%; Score 89; DB 2; Length 23;
Best Local Similarity 77.3%; Pred. No. 0.001;
Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 37 RRVERVLNRRRAAQSRRERKRL 58
   ||:|||||:|||||:|||||:
Db 1 RRERILNRRRAAHQSRERKRL 22

RESULT 5
US-09-148-545-259
/ Sequence 259, Application US/09148545
/ Patent No. 6590075
/ GENERAL INFORMATION:
/ APPLICANT: Rosen et al.
/ TITLE OF INVENTION: 70 Human Secreted Proteins
/ FILE REFERENCE: PZ001P1
/ CURRENT APPLICATION NUMBER: US/09/148,545
/ CURRENT FILING DATE: 1998-09-04
/ EARLIER APPLICATION NUMBER: PCT/US98/04482
/ EARLIER FILING DATE: 1998-03-06
/ EARLIER APPLICATION NUMBER: 60/040,162
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/040,333
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/038,621
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/040,161
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/040,626
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/040,334
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/040,336
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/040,163
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/047,615
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,600
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,597
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,502
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,633
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,583
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,617
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,618
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,503
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,592
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,581
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,584
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,500
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,587
/ EARLIER FILING DATE: 1997-05-23
```

```

, , EARLIER FILING DATE: 1997-08-22
, , EARLIER APPLICATION NUMBER: 60/056,636
, , EARLIER FILING DATE: 1997-08-22
, , EARLIER APPLICATION NUMBER: 60/056,874
, , EARLIER FILING DATE: 1997-08-22
, , EARLIER APPLICATION NUMBER: 60/056,910
, , EARLIER FILING DATE: 1997-08-22
, , EARLIER APPLICATION NUMBER: 60/056,864
, , EARLIER FILING DATE: 1997-08-22
, , EARLIER APPLICATION NUMBER: 60/056,631
, , EARLIER FILING DATE: 1997-08-22
, , EARLIER APPLICATION NUMBER: 60/047,595
, , EARLIER FILING DATE: 1997-05-23
, , EARLIER APPLICATION NUMBER: 60/056,845
, , EARLIER FILING DATE: 1997-08-22
, , EARLIER APPLICATION NUMBER: 60/056,892
, , EARLIER FILING DATE: 1997-08-22
, , EARLIER APPLICATION NUMBER: 60/047,588
, , EARLIER FILING DATE: 1997-05-23
, , EARLIER APPLICATION NUMBER: 60/047,585
, , EARLIER FILING DATE: 1997-05-23
, , EARLIER APPLICATION NUMBER: 60/047,586
, , EARLIER FILING DATE: 1997-05-23
, , EARLIER APPLICATION NUMBER: 60/047,590
, , EARLIER FILING DATE: 1997-05-23
, , EARLIER APPLICATION NUMBER: 60/047,594
, , EARLIER FILING DATE: 1997-05-23
, , EARLIER APPLICATION NUMBER: 60/047,589
, , EARLIER FILING DATE: 1997-05-23
, , EARLIER APPLICATION NUMBER: 60/047,593
, , EARLIER FILING DATE: 1997-05-23
, , EARLIER APPLICATION NUMBER: 60/047,614
, , EARLIER FILING DATE: 1997-05-23
, , EARLIER APPLICATION NUMBER: 60/043,578
, , EARLIER FILING DATE: 1997-04-11
, , EARLIER APPLICATION NUMBER: 60/043,576
, , EARLIER FILING DATE: 1997-04-11
, , EARLIER APPLICATION NUMBER: 60/047,501
, , EARLIER FILING DATE: 1997-05-23
, , EARLIER APPLICATION NUMBER: 60/043,670
, , EARLIER FILING DATE: 1997-04-11
, , EARLIER APPLICATION NUMBER: 60/056,632
, , EARLIER FILING DATE: 1997-08-22
, , EARLIER APPLICATION NUMBER: 60/056,664
, , EARLIER FILING DATE: 1997-08-22
, , EARLIER APPLICATION NUMBER: 60/056,876
, , EARLIER FILING DATE: 1997-08-22
, , EARLIER APPLICATION NUMBER: 60/056,881
, , EARLIER FILING DATE: 1997-08-22
, , EARLIER APPLICATION NUMBER: 60/056,909
, , EARLIER FILING DATE: 1997-08-22
, , EARLIER APPLICATION NUMBER: 60/056,875
, , EARLIER FILING DATE: 1997-08-22
, , EARLIER APPLICATION NUMBER: 60/056,862
, , EARLIER FILING DATE: 1997-08-22
, , EARLIER APPLICATION NUMBER: 60/056,887
, , EARLIER FILING DATE: 1997-08-22
, , EARLIER APPLICATION NUMBER: 60/056,908
, , EARLIER FILING DATE: 1997-08-22
, , EARLIER APPLICATION NUMBER: 60/048,964
, , EARLIER FILING DATE: 1997-06-06
, , EARLIER APPLICATION NUMBER: 60/057,650
, , EARLIER FILING DATE: 1997-09-05
, , EARLIER APPLICATION NUMBER: 60/056,884
, , EARLIER FILING DATE: 1997-08-22
, , NUMBER OF SEQ ID NOS: 260
, , SOFTWARE: PatentIn Ver. 2.0
, , SEQ ID NO 259
, , SEQ LENGTH: 371

```

```

Query Match          27.6%; Score 89; DB 2; Length 371;
Best Local Similarity 42.2%; Pred. No. 0.023;
Matches 27; Conservative 11; Mismatches 16; Indels 10; Gaps 4;

QY      1  EKPPVKRKSQGVLPPEPKTNLPKRKRAKTEDEKEQRRVERVLNRRRAAQSSRRERKRELV 60
Db      126  EKSILLEKE---GLILPE---TLP---LTKTE-EQILAKVRVKIRNKRSQAQESRRKKKVV 175

QY      61  EALE 64
Db      176  GGLE 179

RESULT 6
US-09-621-011-259
; Sequence 259, Application US/09621011
; Patent No. 6878687
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 70 Human Secreted Proteins
; FILE REFERENCE: P2001P1
; CURRENT APPLICATION NUMBER: US/09/621,011
; CURRENT FILING DATE: 2000-07-20
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 280
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 259
; LENGTH: 371
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-011-259

```

Query Match	27.6%;	Score 89;	DB 2;	Length 371;
Best Local Similarity	42.2%;	Pred. No. 0.023;		
Matches	27;	Conservative	11;	Mismatches 16; Indels 10; Gaps 4;
Qy	1	EKPVKVKRSGQVLPEPKTNLPPEPKRAKTEDEKEQRRVVERVLNRRAAQSSRRKRLEV	60	
Db	126	EKSILKE---GLILPE---TLP---LTKTE-EQILKRVRRKIRNKSAQESRRKKVVV	175	
Qy	61	EALF	64	
Db	176	GGLE	179	

```

RESULT 7
US-09-640-211A-1046
; Sequence 1046, Application US/09640211A
; Patent No. 6833446
; GENERAL INFORMATION:
; APPLICANT: Wood, Marion
; APPLICANT: Shenk, Michael A.
; APPLICANT: McGrath, Annette
; APPLICANT: Glenn, Matthew
; TITLE OF INVENTION: Compositions and Methods for the
; FILE OF INVENTION: Modification of Gene Transcription
; FILE REFERENCE: 11000.1021CIU
; CURRENT APPLICATION NUMBER: US/09/640,211A
; CURRENT FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 2368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1046
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-640-211A-1046

```

Query Match 26.6%; Score 86; DB 2; Length 102;
Best Local Similarity 41.5%; Pred. No. 0.012;
Matches 17; Conservative 12; Mismatches 12; Indels 0; Gaps 0;

Db 39 PRKRPRDLSAEKREAPAHNRNRIAAONS RDKRKQOFTSLE 79

[illegible]

```

RESULT 6
US-09-621-011-259
; Sequence 259, Application US/09621011
; Patent No. 6878687
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 70 Human Secreted Proteins
; FILE REFERENCE: P2001P1
; CURRENT APPLICATION NUMBER: US/09/621,011
; CURRENT FILING DATE: 2000-07-20
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 280
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 259
; LENGTH: 371
; TYPE: prt
; ORGANISM: Homo sapiens
US-09-621-011-259

```

Query Match	27.6%;	Score 89;	DB 2;	Length 371;
Best Local Similarity	42.2%;	Pred. No. 0.023;		
Matches	27;	Conservative	11;	Mismatches 16; Indels 10; Gaps 4;
Qy	1	EKPVKVKRSGQVLPEPKTNLPPEPKRAKTEDEKEQRRVVERVLNRRAAQSSRRKRLEV	60	
Db	126	EKSILKE---GLILPE---TLP---LTKTE-EQILKRVRRKIRNKSAQESRRKKVVV	175	
Qy	61	EALF	64	
Db	176	GGLE	179	

```

RESULT 7
US-09-640-211A-1046
; Sequence 1046, Application US/09640211A
; Patent No. 6833446
; GENERAL INFORMATION:
; APPLICANT: Wood, Marion
; APPLICANT: Shenk, Michael A.
; APPLICANT: McGrath, Annette
; APPLICANT: Glenn, Matthew
; TITLE OF INVENTION: Compositions and Methods for the
; FILE OF INVENTION: Modification of Gene Transcription
; FILE REFERENCE: 11000.1021CIU
; CURRENT APPLICATION NUMBER: US/09/640,211A
; CURRENT FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 2368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1046
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-640-211A-1046

```

Query Match 26.6%; Score 86; DB 2; Length 102;
Best Local Similarity 41.5%; Pred. No. 0.012;
Matches 17; Conservative 12; Mismatches 12; Indels 0; Gaps 0;

```

RESULT 8
US-09-216-393B-81
; Sequence 81, Application US/09216393B
; Patent No. 6514694
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC
; FILE REFERENCE: TX-1-C2
; CURRENT APPLICATION NUMBER: US/09/216,393B
; CURRENT FILING DATE: 1998-12-18
; PRIOR APPLICATION NUMBER: 08/994,825
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 366
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 81
; LENGTH: 611
; TYPE: PRT
; ORGANISM: Toxoplasma gondii
US-09-216-393B-81

```

```

Query Match      25.5%; Score 82.5; DB 2; Length 611;
Best Local Similarity 35.7%; Pred. NO. 0.22;
Matches 20; Conservative 10; Mismatches 23; Indels 3; Gaps 1;

Qy 12 GQVLPEPKT---NLPPRKAKTEDEKQRRVERVLNRRRAAQSSRRKRKLEVALE 64
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 277 GMSPTADTTRHDAEERRRRAAEERKQREERERREVRVEKEKREQEEERE 332

```

RESULT 9
US-09-489-039A-13074
; Sequence 13074, Application US/09489039A
; Patent No. 6610836

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

```

; FILE REFERENCE: 2705-2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13074

```

US-09-489-039A-13074

Qy 2 KXPVKRKSNGV---LPB-----PKTNLPKRKRAKTEDEKEQRVERVLNRR 47
 ||| : || | - : || : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 127 KPSSRLSRROAKRLPKKKPLKPPRKRPPLRPPPRPOPLKRLQOBLQLTKRPRR 186

Qy 48 AAQSSRRKR 57
187 OOKP RRRRK 196

RESULT 10
US-09-247-155-113
; Sequence 113, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-
; APPLICANT: Duclert, Aymeric

APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET.021A
; CURRENT APPLICATION NUMBER: US/09/247,155A
; CURRENT FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/099,273
; EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 113
; LENGTH: 395;
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -310...-1
US-09-247-155-113

Query Match 24.9%; Score 80.5; DB 2; Length 395;
Best Local Similarity 27.6%; Pred. No. 0.23; Mismatches 13; Indels 23; Gaps 1;
Matches 21; Conservative 13

QY 12 GOVLPEPTNLPPRKRAKTEDEKEQ-----RRRVRLNRRA 48
DB 171 GTVAVPCTTLPCQTLFTDEKRLGQGVSLPSHLPLTKAERVLKVKRKIRNKQS 230
QY 49 AQSSRRKRLEVEALE 64
DB 231 AQSSRRRKKEYIDGLE 246

RESULT 11
US-09-513-999C-14
; Sequence 14, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59 US2 REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 14
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -310...-1
; OTHER INFORMATION: score 4.4
; OTHER INFORMATION: seq VLLILFSLALIL/PS
US-09-513-999C-14

Query Match 24.9%; Score 80.5; DB 2; Length 395;
Best Local Similarity 27.6%; Pred. No. 0.23; Mismatches 13; Indels 23; Gaps 1;
Matches 21; Conservative 13

QY 12 GOVLPEPTNLPPRKRAKTEDEKEQ-----RRRVRLNRRA 48
DB 171 GTVAVPCTTLPCQTLFTDEKRLGQGVSLPSHLPLTKAERVLKVKRKIRNKQS 230

QY 49 AQSSRRKRLEVEALE 64
DB 231 AQSSRRRKKEYIDGLE 246
RESULT 12
US-09-471-276-14
; Sequence 14, Application US/09471276
; Patent No. 6822072
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6822072
; FILE REFERENCE: GENSET.025CPI
; CURRENT APPLICATION NUMBER: US/09/471,276
; CURRENT FILING DATE: 1999-12-21
; EARLIER APPLICATION NUMBER: 09/057,719
; EARLIER FILING DATE: 1998-04-09
; EARLIER APPLICATION NUMBER: 09/069,047
; EARLIER FILING DATE: 1998-04-28
; EARLIER APPLICATION NUMBER: PCT/IB99/00712
; EARLIER FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 1622
; SOFTWARE: Patent.pm
; SEQ ID NO 14
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -310...-1
; OTHER INFORMATION: score 4.4
; OTHER INFORMATION: seq VLLILFSLALIL/PS
US-09-471-276-14

Query Match 24.9%; Score 80.5; DB 2; Length 395;
Best Local Similarity 27.6%; Pred. No. 0.23; Mismatches 13; Indels 23; Gaps 1;
Matches 21; Conservative 13

QY 12 GOVLPEPTNLPPRKRAKTEDEKEQ-----RRRVRLNRRA 48
DB 171 GTVAVPCTTLPCQTLFTDEKRLGQGVSLPSHLPLTKAERVLKVKRKIRNKQS 230
QY 49 AQSSRRKRLEVEALE 64
DB 231 AQSSRRRKKEYIDGLE 246

RESULT 13
US-09-903-190-113
; Sequence 113, Application US/09903190
; Patent No. 6936592
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET.021A
; CURRENT APPLICATION NUMBER: US/09/903,190
; CURRENT FILING DATE: 2001-07-11
; PRIOR FILING DATE: 1999-02-09
; PRIOR FILING DATE: 1999-02-09
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/074,121
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-02-09
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/081,563
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/096,116
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-10
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/099,273
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182

```

; SOFTWARE: Patent.pm
; SEQ ID NO 113
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -310...-1
US-09-903-190-113

Query Match      24.9%; Score 80.5; DB 2; Length 395;
Best Local Similarity 27.6%; Pred. No. 0.23;
Matches 21; Conservative 13; Mismatches 19; Indels 23; Gaps 1;

QY 12 GOVLPEPKTNLPKRKTEDEKEQ-----RRVERVLRNRA 48
Db 171 GTVAPVCTLLPCQTLTDEKRLGQVSLPDLTKAERVLKVRKIRNKQS 230
QY 49 AQSSRRKRLEVEAL 64
Db 231 AQDSRRRKKEYIDGLE 246

RESULT 14
US-09-107-433-3674
; Sequence 3674, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3674:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 63 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...63
; SEQUENCE DESCRIPTION: SEQ ID NO: 3674:

```

US-09-107-433-3674

Query Match 24.8%; Score 80; DB 2; Length 63;
Best Local Similarity 32.2%; Pred. No. 0.034; 24; Indels 4; Gaps 1;
Matches 19; Conservative 12; Mismatches 24; Indels 4; Gaps 1;

QY 1 EKPKVKRKRGQVLPKPNLPKRKTEDEKEQRRVERVLRNRAAQSSRRKRLE 59
Db 8 KKK 62

RESULT 15

US-08-796-899-29
; Sequence 29, Application US/08796899
; Patent No. 6160202
; GENERAL INFORMATION:
; APPLICANT: BUSTOS, Mauricio M
; APPLICANT: CHERN, Maw-Sheng
; TITLE OF INVENTION: MODIFICATION OF SEED CROPS WITH
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/796,899
; FILING DATE: 06-FEB-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/319,544
; FILING DATE: 07-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 028754-005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 551 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-796-899-29

Query Match 24.1%; Score 78; DB 2; Length 551;
Best Local Similarity 39.6%; Pred. No. 0.64; 18; Indels 2; Gaps 1;
Matches 19; Conservative 9; Mismatches 18; Indels 2; Gaps 1;

QY 16 PEKTNLPKRKTEDEKEQRRVERVLRNRAAQSSRRKRLEVEAL 63
Db 287 PQPCAAPP--EAWLQNERELKRRKQSNRESARRSLRKOATEEL 332

Search completed: November 22, 2005, 15:50:42
Job time : 34.5 secs

Result No.	Score	Query			DB	ID	Description
		Match	Length				
1	175	54.2	238	2	S78571	transcription fact	
2	106	32.8	168	2	T50922	bzip protein HV5 [
3	91	28.2	322	2	T08592	TGACG-motif-bindin	
4	91	28.2	326	2	T08591	TGACG-motif-bindin	
5	90.5	28.0	688	2	T32750	hypothetical prote	
6	88	27.2	322	2	T12093	TGACG-motif bindin	
7	87.5	27.1	176	2	B90087	hypothetical prote	
8	87	26.9	600	2	T00759	hypothetical prote	
9	86	26.6	360	2	T03373	probable G-box bin	
10	81	25.1	445	2	T50972	probable zutotin [i	
11	80.5	24.9	486	2	JC4028	activating transcr	
12	79.5	24.6	2052	2	T18519	myosin X - bovine	
13	79	24.5	381	2	S26812	transcription fact	
14	79	24.5	672	2	T21469	hypothetical prote	
15	78.5	24.3	232	2	S42392	G-box-binding prot	
16	78.5	24.3	242	2	S05453	transcription fact	
17	78	24.1	424	2	T10985	regulator protein	
18	77.5	24.0	246	2	T12585	Dc3 promoter-bindi	
19	77.5	24.0	483	2	S12741	transcription fact	
20	77.5	24.0	505	1	S05380	transcription fact	
21	77	23.8	1089	2	T36663	protein kinase, tr	
22	76.5	23.7	144	2	T14796	hypothetical prote	
23	76.5	23.7	313	2	A34785	DNA-binding protei	
24	76.5	23.7	351	2	A45377	transcription fact	
25	76.5	23.7	358	2	C42026	cyclic AMP respons	
26	76.5	23.7	389	1	A39429	cAMP response elem	
27	76.5	23.7	448	2	A42026	cAMP response elem	
28	76.5	23.7	456	2	B42026	cyclic AMP respons	
29	76	23.5	502	2	T20130	hypothetical prote	

C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004
C;Accession: T50922
R;Oyama, T.; Shimura, Y.; Okada, K.
submitted to the EMBL Data Library, July 1997
A;Description: The Arabidopsis HY5 gene encodes a bZIP protein that regulates stimulus-induced transcriptional activation of the HY5 promoter
A;Reference number: Z25271
A;Accession: T50922
A;Status: preliminary; translated from GB/EMBL/DDBB
A;Molecule type: mRNA
A;Residues: 1-168 <OVA>
A;Cross-references: UNIPROT:Q24646; UNIPARC:UPI000012CFCA; EMBL:AB005295; PIDN:BAA21116
A;Experimental source: Landsberg erecta

Query Match 32.8%; Score 106; DB 2; Length 168;
Best Local Similarity 43.4%; Pred.No. 0.0078;
Matches 23; Conservative 10; Mismatches 20; Indels 0; Gaps 0;

QY 12 GOVLPEKTNLPKRAKTEDEKEQRVERVLRRAAQSSRRKRLEVEALE 64
Db 65 GOERTATGVSQKRGRTPAEKENRLKLLNRVSAQAARERKKAYLSELE 117

RESULT 3

T08592

GCAG-motif-binding protein STF2 - soybean
C;Species: Glycine max (soybean)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 31-Dec-2004
C;Accession: T08592
R;Cheong, Y.H.; Yoo, C.M.; Park, J.M.; Ryu, G.R.; Goekjian, V.H.; Nagao, R.T.; Key, J.L.
submitted to the EMBL Data Library, September 1995
A;Description: STF1 is a novel GCAG-binding factor with a zinc-finger motif and a bZIP
A;Reference number: Z16445
A;Accession: T08592
A;Status: preliminary; translated from GB/EMBL/DDBB
A;Molecule type: mRNA
A;Residues: 1-322 <CHE>
A;Cross-references: UNIPROT:Q39896; UNIPARC:UPI00000A61C8; EMBL:L28004; NID:g986966; PFI
A;Experimental source: strain Williams; hypocotyl
C;Genetics:
A;Gene: STF2
C;Superfamily: GCAG-motif-binding transcription factor

Query Match 28.2%; Score 91; DB 2; Length 322;
Best Local Similarity 45.0%; Pred.No. 0.31;
Matches 18; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

QY 25 RKAKTEDEKEQRVERVLRRAAQSSRRKRLEVEALE 64
Db 234 KKGRSPADKESKRLKLLNRVSAQAARERKKAYLIDLE 273

RESULT 4

T08591

GCAG-motif binding protein STF1 - soybean
C;Species: Glycine max (soybean)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 31-Dec-2004
C;Accession: T08591
R;Cheong, Y.H.; Yoo, C.M.; Park, J.M.; Ryu, G.R.; Goekjian, V.H.; Nagao, R.T.; Key, J.L.
submitted to the EMBL Data Library, September 1995
A;Description: STF1 is a novel GCAG-binding factor with a zinc-finger motif and a bZIP
A;Reference number: Z16445
A;Accession: T08591
A;Status: preliminary; translated from GB/EMBL/DDBB
A;Molecule type: mRNA
A;Residues: 1-326 <CHE>
A;Cross-references: UNIPROT:Q39895; UNIPARC:UPI00000A6233; EMBL:L28003; NID:g2934883; PFI
A;Experimental source: strain Williams; hypocotyl
C;Superfamily: GCAG-motif-binding transcription factor

Query Match 28.2%; Score 91; DB 2; Length 326;
Best Local Similarity 45.0%; Pred.No. 0.32;
Matches 18; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 31-Dec-2004

C/Accession: T03373
R:Nakagawa, H.; Ohmiya, K.; Hattori, T.
Plant J. 9, 217-227, 1996

A>Title: A rice bZIP protein, designated OSB28, is rapidly induced by abscisic acid.

A/Reference number: Z14906; MUID:96417817; PMID:8820608

A/Accession: T03373

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-360 <NAK>

A/Cross-references: UNIPROT:Q40645; UNIPARC:UPI00000A8257; EMBL:U42208; NID:g1147631; PII

A/Experimental source: cv. Nipponbare

C/Superfamily: BZIP G-box-binding protein; fos/jun DNA-binding domain homology

Query Match 26.6%; Score 86; DB 2; Length 360;
Best Local Similarity 37.7%; Pred. No. 0.98;
Matches 23; Conservative 9; Mismatches 23; Indels 6; Gaps 1;

Qy 3 KPVKKRSGOVLPPEKTNLPKRKAKTEDEKEQRRVLENNRRAAQSSRRKRLEVEA 62
Db 199 KPDVSTASDFRIATPVTETP-----TKDDKSKRRRKQSNRESARRLRKQAEET 252

Qy 63 L 63
Db 253 L 253

RESULT 10

T50972

Probable zeatin [imported] - Neurospora crassa

N/Alternate names: protein B24P7.270

C/Species: Neurospora crassa

C/Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004

C/Accession: T50972

R/Schulte, U.; Agn, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, submitted to the Protein Sequence Database, July 2000

A/Reference number: Z25286

A/Accession: T50972

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-445 <SCH>

A/Cross-references: UNIPROT:Q9P3Q8; UNIPARC:UPI000006B021; EMBL:ALJ389890; GSPDB:GN00116;

A/Experimental source: BAC clone B24P7; strain OR74A

C/Genetics:

A/Gene: NCSP:B24P7.270

A/Map position: 6

A/Introns: 98/3

Query Match 25.1%; Score 81; DB 2; Length 445;
Best Local Similarity 35.2%; Pred. No. 3.3;
Matches 19; Conservative 11; Mismatches 16; Indels 8; Gaps 1;

Qy 19 KTNLPKRKAKTEDEKQRRV-----ERVLENNRRAAQSSRRKRLEVEALE 64
Db 260 KKNLNRKKRAEDNARLRKLLDCCSAADERIKKFRQENAAKNKKRLEKAAE 313

RESULT 11

JC4028

activating transcription factor 2 - African clawed frog

N/Alternate names: cyclic AMP-response element-binding protein

C/Species: Xenopus laevis (African clawed frog)

C/Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 09-Jul-2004

C/Accession: JC4028

R/Villarreal, X.C.; Richter, J.D.

Gene 153, 225-229, 1995

A/Title: Analysis of ATF2 gene expression during early Xenopus laevis developme nt.

A/Reference number: JC4028; MUID:95180723; PMID:7875593

A/Accession: JC4028

A/Molecule type: mRNA

A/Residues: 1-486 <VIL>

A/Cross-references: UNIPROT:Q91576; UNIPARC:UPI00000FB2FD; GB:U16158; NID:g987779; PIDs:1

C/Comment: This protein is a sequence-specific DNA-binding protein that mediates transcr

C; Superfamily: cAMP response element-binding protein 1; fos/jun DNA-binding domain homolog
C; Keywords: leucine zipper; phosphoprotein; transcription regulation
F; 37-398/Region: leucine zipper motif
F; 329-369/Domain: fos/jun DNA-binding domain homology <FJD>
F; 332-364/Region: basic
F; 82/Binding site: phosphate (Ser) (covalent) (by protein kinase A) #status predicted
F; 102/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status predicted

Query Match 24.9%; Score 80.5; DB 2; Length 486;
Best Local Similarity 33.3%; Pred. No. 4;
Matches 19; Conservative 16; Mismatches 21; Indels 1; Gaps 1;

Qy 8 RKSNGVQLPEPKTNLPKRKAKTDESKQRVERVLNRRAAQSSRRKRLEVEALE 64
| | | : : : || : : || : : || : : || : : || : : || : : ||
Db 308 RPSPAQPPTQTOSTGRRRAANED-PDEKRSKIIOENRAAAACRCRKKRVVVQSLE 363

RESULT 12
Tl8519
myosin X - bovine
C; Species: Bos primigenius taurus (cattle)
C; Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C; Accession: Tl8519
R; Corey, D.P.; Derfler, B.H.; Solc, C.K.; Duyk, G.M.; Cheney, R.E.
submitted to the EMBL Data Library, April 1996
A; Description: Cloning and expression of myosin X, a novel unconventional myosin with pl
A; Reference number: Z18942
A; Accession: Tl8519
A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: mRNA
A; Residues: 1-2052 <COR>
A; Cross-references: UNIPROT:P79114; UNIPARC:UPI000012PABC; EMBL:U55042; NID:g1755048; PI
A; Experimental source: aorta
C; Keywords: nucleotide binding; P-loop
F; 66-727/Domain: myosin motor domain homology <MMO>
F; 157-164/Region: nucleotide-binding motif A (P-loop)

Query Match 24.6%; Score 79.5; DB 2; Length 2052;
Best Local Similarity 35.5%; Pred. No. 18;
Matches 22; Conservative 10; Mismatches 25; Indels 5; Gaps 1;

Qy 8 RKSNGVQLPEPKTNLPKRK-----AKTEDEKQRVERVLNRRAAQSSRRKRLEVEA 62
| : | : | : ||| | : | : | : ||| | : | : | : ||| | : | : | : |||
Db 805 RVYVROLLAEKRAEEKEEKEKREEREREEREREAELRAQQEEAAKQRELEA 864

Qy 63 LE 64
| :
Db 865 LQ 866

RESULT 13
S26812
transcription factor ATF-4 - mouse
N; Alternate names: activating transcription factor 4; TAXREB67 homolog; transcription fa
C; Species: Mus musculus (house mouse)
C; Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Oct-2004
C; Accession: S26812; S18719; C46132; A58861; A47443
R; Pruitt, S.C.
submitted to the EMBL Data Library, August 1991
A; Reference number: S26812
A; Accession: S26812
A; Molecule type: DNA
A; Residues: 1-381 <PRU>
A; Cross-references: UNIPROT:Q61328; UNIPARC:UPI000000283A5; EMBL:X61507; NID:g50049; PIDN
R; Mielnicki, L.M.; Pruitt, S.C.
Nucleic Acids Res. 19, 6332, 1991
A; Title: Isolation and nucleotide sequence of a murine cDNA homologous to human activati
A; Reference number: S18719; MUID:Z9206493; PMID:1956797
A; Accession: S18719
A; Molecule type: DNA
A; Residues: 1-57,59-288,'R',290-381 <MIE>
A; Cross-references: UNIPARC:UPI0000179B78; EMBL:X61507
R; Chevray, P.M.; Nathans, D.

RESULT 15

S42392
G-box-binding protein - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 31-Dec-2004
C:Accession: S42392
R:Meier, I.; Gruissem, W.
Nucleic Acids Res. 22, 470-478, 1994
A:Title: Novel conserved sequence motifs in plant G-box binding proteins and implication
A:Reference number: S42392; MUID:94173701; PMID:8127687
A:Accession: S42392
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-232 <ME1>
A:Cross-references: UNIPROT:Q43507; UNIPARC:UPI00000ACF8F; EMBL:X74941; NID:g456750; PID
C:Superfamily: BZIP G-box-binding protein; fos/jun DNA-binding domain homology
F:130-170/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 24.3%; Score 78.5; DB 2; Length 232;
Best Local Similarity 35.0%; Pred. No. 3.1;
Matches 21; Conservative 10; Mismatches 24; Indels 5; Gaps 1;
Oy 5 VKKRKSWGQVLPEPTNLPPRKRAKTEDEKEQRRVERVLNRRAAQSSRRKRLEVEALE 64
Db 110 IKWRSNQSGVSPAGWG-----REWIQDERELKQKQKQSNRESARRSLRKQACEELQ 164

Search completed: November 22, 2005, 15:49:29
Job time : 26 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2005, 15:31:54 ; Search time 159 Seconds
(without alignments)
176.857 Million cell updates/sec

Title: US-10-663-450-5_COPY_84_147

Perfect score: 323

Sequence: 1 EKKPVKKRSGQVLPEPKT.....NRRAAQSSRRKRLEVEALS 64

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_21.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	323	100.0	64	5 AAE15373	Aae15373 Trichoder
2	323	100.0	450	4 AAB82975	Aab82975 Trichoder
3	323	100.0	451	5 AAE15371	Aae15371 Trichoder
4	323	100.0	451	8 ADS12801	Adsi12801 Trichoder
5	323	100.0	451	8 ADS12804	Adsi12804 Trichoder
6	307	95.0	409	7 ABO43144	Abo43144 A. thalia
7	307	95.0	409	7 ADB31925	Adb31925 Plant (A.
8	307	95.0	409	8 ADO02271	Ado02271 Thalecres
9	292	90.4	342	4 AAB82977	Aab82977 Aspergill
10	292	90.4	342	5 AAE15381	Aae15381 Aspergill
11	292	90.4	342	8 ADS12818	Adsi12818 Aspergill
12	292	90.4	366	5 AAE15379	Aae15379 Aspergill
13	292	90.4	386	8 ADS12815	Adsi12815 Aspergill
14	288	89.2	64	5 AAE15374	Aae15374 Aspergill
15	288	89.2	349	8 ADS12805	Adsi12805 Aspergill
16	288	89.2	349	8 ADS12803	Adsi12803 Aspergill
17	288	89.2	349	8 ADS12863	Adsi12863 Aspergill
18	288	89.2	350	4 AAB82976	Aab82976 Aspergill
19	288	89.2	350	5 AAE15372	Aae15372 Aspergill
20	175	54.2	68	5 AAE15382	Aae15382 Yeast HAC
21	175	54.2	68	8 ADS12859	Adsi12859 Saccharom
22	175	54.2	84	5 ABP02534	Abp02534 Human ORF
23	175	54.2	200	8 ADS43437	Ads43437 Bacterial
24	175	54.2	230	2 AAW53806	Aaw53806 Transcrip

25	175	54.2	230	8 ADT87049	Adt87049 Yeast Str
26	175	54.2	238	2 AAW53807	Aaw53807 Transcrip
27	108	33.4	192	4 AAB82614	Aab82614 Maize roo
28	108	33.4	192	4 AAB82615	Aab82615 Maize roo
29	108	33.4	192	4 AAB82616	Aab82616 Maize roo
30	108	33.4	192	4 AAG66525	Aag66525 Maize roo
31	108	33.4	192	4 AAG66526	Aag66526 Maize roo
32	106	32.8	168	5 AAU93013	Aau93013 Arabidops
33	106	32.8	168	7 ADD30174	Add30174 Plant yie
34	106	32.8	168	8 ADI43893	Adi43893 Plant tra
35	106	32.8	211	3 AAG08861	Aag08861 Arabidops
36	103	31.9	170	8 ADM48147	Adm48147 Polypteti
37	101	31.3	672	8 ADX95805	Adx95805 Plant ful
38	100.5	31.1	185	9 ADM17162	Adm17162 Eucalyptu
39	99	30.7	143	3 AAB33151	Aab33151 Pinus rad
40	99	30.7	163	9 ADM17580	Adm17580 Pinus rad
41	94	29.1	646	7 ABM86095	Abm86095 Rice abio
42	93	28.8	120	3 AAG27808	Aag27808 Arabidops
43	93	28.8	135	3 AAG07181	Aag07181 Arabidops
44	93	28.8	149	3 AAG07180	Aag07180 Arabidops
45	93	28.8	149	3 AAG27807	Aag27807 Arabidops

ALIGNMENTS

RESULT 1

AAE15373

ID AAE15373 standard; protein; 64 AA.

AC AAE15373;

XX

DT 29-AUG-2003 (revised)

DT 07-MAR-2002 (first entry)

XX

DE Trichoderma reesei HAC1 protein DNA binding domain.

XX

KW Heterologous protein secretion; unfolded protein response; UPR; lipase;

KW cellulase; carbohydrase; industry; purification; DNA binding domain;

KW HAC1 protein.

OS Hypocrea jecorina.

XX

PN US2001034045-A1.

XX

PD 25-OCT-2001.

XX

PF 23-MAR-2001; 2001US-00816277.

XX

PR 24-MAR-2000; 2000US-00534692.

XX

PA (GEMV) GENENCOR INT INC.

XX

PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;

XX

DR WPI; 2002-033728/04.

XX

PT Increasing secretion of heterologous proteins e.g. lipase and cellulase in eukaryotic cells useful in industry to increase production and facilitate purification, by inducing an elevated unfolded protein response.

PT

PS Example 3; Fig 10; 56pp; English.

XX

CC The present invention relates to methods for increasing the secretion of heterologous protein in eukaryotic cells by inducing an elevated UPR by protein response (UPR). The method involves inducing the elevated UPR by increasing the presence of proteins such as HAC1, HACA, PTC2 or IRE1 in cells. The method and sequences are useful for increasing the secretion of heterologous proteins (e.g. lipase, cellulase, carbohydrase) in eukaryotic cells useful in industry to increase protein yields and to facilitate purification. The present sequence is Trichoderma reesei HAC1 protein DNA binding domain. (Updated on 29-AUG-2003 to standardise OS

```
CC field)
XX
SQ Sequence 64 AA;
    Query Match      100.0%; Score 323; DB 5; Length 64;
    Best Local Similarity 100.0%; Pred. No. 1.5e-29;
    Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKKPVKKRSWGQVLPEPKTNLPPrKRKTEDEKEQRRVERVLNRRAAQSRRKRLEV 60
    |||||
DB 1 EKKPVKKRSWGQVLPEPKTNLPPrKRKTEDEKEQRRVERVLNRRAAQSRRKRLEV 60
    |||||
QY 61 EALE 64
    ||||
DB 61 EALE 64

RESULT 2
AAB82975
ID AAB82975 standard; protein; 450 AA.
XX
AC AAB82975;
XX
DT 11-SEP-2003 (revised)
DT 21-DEC-2001 (first entry)
XX
DE Trichoderma reesei HAC1, involved in unfolded protein response.
XX
KW HAC1; transcription factor; unfolded protein response; protein secretion.
XX
OS Hypocrea jecorina.
XX
FH Key Location/Qualifiers
FT Domain 84..147
FT /label= DNA binding domain
XX
PN WO200172783-A2.
XX
PD 04-OCT-2001.
XX
PF 23-MAR-2001; 2001WO-US009401.
XX
PR 24-MAR-2000; 2000US-00534692.
XX
PA (GENV ) GENENCOR INT INC.
XX
PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX
DR WPI; 2001-626252/72.
DR N-PSDB; AAB26931.
XX
PT Increasing secretion of heterologous proteins e.g. lipase and cellulase
PT in eukaryotic cells useful in industry to increase production and
PT facilitate purification, by inducing an elevated unfolded protein
PT response.
XX
PS Claim 54; Fig 7A-B; 89pp; English.
XX
CC The present sequence is that of the HAC1 protein of Trichoderma reesei,
CC as deduced from the newly isolated HAC1 gene (see AAB26931). HAC1 protein
CC is a transcription factor involved in the unfolded protein response
CC (UPR). The invention provides methods for increasing the secretion of a
CC heterologous protein in a cell by inducing an elevated UPR. This can be
CC achieved by modulating the activity of HAC1, PTC2 or IRE1 in the cell,
CC e.g. by gene overexpression. The cell from which the protein is secreted
CC can be any cell having an UPR, such as mammalian cells, insect cells,
CC yeast and filamentous fungi. The protein of interest can be any secreted
CC protein such as a therapeutic protein or an industrial enzyme, e.g.
CC lipase, cellulase, endoglucanase-H, protease, carbohydratase, reductase,
CC oxidase, isomerase, transferase, kinase, phosphatase, alpha-amylase,
CC glucosylase, lignocellulose hemicellulase, pectinase and ligninase
CC (claimed). (Updated on 11-SEP-2003 to standardise OS field)
XX

SQ Sequence 450 AA;
    Query Match      100.0%; Score 323; DB 4; Length 450;
    Best Local Similarity 100.0%; Pred. No. 1.3e-28;
    Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKKPVKKRSWGQVLPEPKTNLPPrKRKTEDEKEQRRVERVLNRRAAQSRRKRLEV 60
    |||||
DB 84 EKKPVKKRSWGQVLPEPKTNLPPrKRKTEDEKEQRRVERVLNRRAAQSRRKRLEV 143
    |||||
QY 61 EALE 64
    ||||
DB 144 EALE 147

RESULT 3
AAE15371
ID AAE15371 standard; protein; 451 AA.
XX
AC AAE15371;
XX
DT 29-AUG-2003 (revised)
DT 07-MAR-2002 (first entry)
XX
DE Trichoderma reesei HAC1 protein.
XX
KW Heterologous protein secretion; unfolded protein response; UPR; lipase;
KW cellulase; carbohydrase; industry; purification; HAC1 protein.
XX
OS Hypocrea jecorina.
XX
FH Key Location/Qualifiers
FT Binding-site 84..147
FT /label= DNA-binding_domain
XX
PN US2001034045-A1.
XX
PD 25-OCT-2001.
XX
PF 23-MAR-2001; 2001US-00816277.
XX
PR 24-MAR-2000; 2000US-00534692.
XX
PA (GENV ) GENENCOR INT INC.
XX
PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX
DR WPI; 2002-033728/04.
DR N-PSDB; AAD24595.
XX
PT Increasing secretion of heterologous proteins e.g. lipase and cellulase
PT in eukaryotic cells useful in industry to increase production and
PT facilitate purification, by inducing an elevated unfolded protein
PT response.
XX
PS Claim 39; Fig 7; 56pp; English.
XX
CC The present invention relates to methods for increasing the secretion of
CC heterologous protein in eukaryotic cells by inducing an elevated unfolded
CC protein response (UPR). The method involves inducing the elevated UPR by
CC increasing the presence of proteins such as HAC1, HACA, PTC2 or IRE1 in
CC cells. The method and sequences are useful for increasing the secretion
CC of heterologous proteins (e.g. lipase, cellulase, carbohydrase) in
CC eukaryotic cells useful in industry to increase protein yields and to
CC facilitate purification. The present sequence is Trichoderma reesei HAC1
CC protein. (Updated on 29-AUG-2003 to standardise OS field)
XX
SQ Sequence 451 AA;
    Query Match      100.0%; Score 323; DB 5; Length 451;
    Best Local Similarity 100.0%; Pred. No. 1.3e-28;
    Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 EKKPVKRSWGQVLPEPKTNLPKRKAKTEDEKEQRRVERVLNRRRAAQSRRKRLEV 60
|||||
Db 84 EKKPVKRSWGQVLPEPKTNLPKRKAKTEDEKEQRRVERVLNRRRAAQSRRKRLEV 143
QY 61 EALE 64
|||||
Db 144 EALE 147

RESULT 4
ADS12801
ID ADS12801 standard; protein; 451 AA.
AC ADS12801;
XX
XX
DT 16-DEC-2004 (first entry)
DE Trichoderma reesei hac1 chaperone and foldase.
XX
XX unfolded protein response; UPR; HAC1; PTC2;
KW unfolded protein response modulator; enzyme production; hac1; chaperone;
KW foldase; enzyme.
XX
XX Hypocrea jecorina.
XX
PN US2004186070-A1.
XX
XX
PD 23-SEP-2004.
XX
XX 15-SEP-2003; 2003US-00663450.
XX
XX 24-MAR-2000; 2000US-00534692.
PR 23-MAR-2001; 2001US-00816277.
XX
XX (GEMV) GENENCOR INT INC.
XX
XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX WPI; 2004-707924/69.
DR N-PSDB; ADS12800.
XX
XX Increasing the secretion of a heterologous protein, such as a therapeutic
PT or an industrial enzyme, in genetically modified eukaryotic cells by
PT inducing an elevated unfolded protein response (UPR).
XX
XX Example 3; SEQ ID NO 2; 83pp; English.
XX
XX The invention describes a method of increasing the secretion of a
CC heterologous protein in a eukaryotic cell, comprising inducing an
CC elevated unfolded protein response (UPR). Also described are: an isolated
CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
CC unfolded protein response and has less than 50% similarity to yeast HAC1
CC protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
CC protein induces unfolded protein response and where the HAC1 protein
CC comprises a DNA binding region that has greater than 70% similarity to
CC the DNA binding region of filamentous fungi HAC1 protein; a protein
CC having unfolded protein response inducing activity and having greater
CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
CC respectively) as given in the specification; a protein having an amino
CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
CC response and has at least 70% similarity to a fully defined amino acid
CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
CC specification; an isolated nucleic acid encoding a IRE1 protein that
CC modulates unfolded protein response and has at least 60% similarity to a
CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
CC having unfolded protein response modulating activity and having greater
CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
CC heterologous nucleic acid encoding a protein having unfolded protein
CC response modulating activity and a heterologous nucleic acid encoding a
CC protein of interest to be secreted. The methods and compositions of
CC genetically manipulating cells to have an elevated unfolded protein

CC response (UPR) resulting in an increased capacity to produce secreted
CC proteins, are useful in e.g. production of therapeutic or industrial
CC enzymes. This is the amino acid sequence of Trichoderma reesei HAC1
CC chaperone and foldase.
XX
XX Sequence 451 AA;
SQ

Query Match 100.0%; Score 323; DB 8; Length 451;
Best Local Similarity 100.0%; Pred. No. 1.3e-28;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX

QY 1 EKKPVKRSWGQVLPEPKTNLPKRKAKTEDEKEQRRVERVLNRRRAAQSRRKRLEV 60
|||||
Db 84 EKKPVKRSWGQVLPEPKTNLPKRKAKTEDEKEQRRVERVLNRRRAAQSRRKRLEV 143
QY 61 EALE 64
|||||
Db 144 EALE 147

RESULT 5
ADS12804
ID ADS12804 standard; protein; 451 AA.
XX
XX ADS12804;
AC
XX
DT 16-DEC-2004 (first entry)
DE Trichoderma reesei hac1 DNA binding domain.
XX
XX
XX
XX unfolded protein response; UPR; HAC1; PTC2;
KW unfolded protein response modulator; enzyme production; hac1; chaperone;
KW foldase; DNA binding domain.
XX
XX
XX Hypocrea jecorina.
XX
XX US2004186070-A1.
XX
XX 23-SEP-2004.
XX
XX 15-SEP-2003; 2003US-00663450.
XX
XX 24-MAR-2000; 2000US-00534692.
PR 23-MAR-2001; 2001US-00816277.
XX
XX (GEMV) GENENCOR INT INC.
XX
XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX WPI; 2004-707924/69.
XX
XX Increasing the secretion of a heterologous protein, such as a therapeutic
PT or an industrial enzyme, in genetically modified eukaryotic cells by
PT inducing an elevated unfolded protein response (UPR).
XX
XX Example 3; SEQ ID NO 5; 83pp; English.
XX
XX The invention describes a method of increasing the secretion of a
CC heterologous protein in a eukaryotic cell, comprising inducing an
CC elevated unfolded protein response (UPR). Also described are: an isolated
CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
CC unfolded protein response and has less than 50% similarity to yeast HAC1
CC protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
CC protein induces unfolded protein response and where the HAC1 protein
CC comprises a DNA binding region that has greater than 70% similarity to
CC the DNA binding region of filamentous fungi HAC1 protein; a protein
CC having unfolded protein response inducing activity and having greater
CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
CC respectively) as given in the specification; a protein having an amino
CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
CC response and has at least 70% similarity to a fully defined amino acid
CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
CC specification; an isolated nucleic acid encoding a IRE1 protein that
CC modulates unfolded protein response and has at least 60% similarity to a
CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
CC having unfolded protein response modulating activity and having greater
CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
CC heterologous nucleic acid encoding a protein having unfolded protein
CC response modulating activity and a heterologous nucleic acid encoding a
CC protein of interest to be secreted. The methods and compositions of
CC genetically manipulating cells to have an elevated unfolded protein

CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
CC specification; an isolated nucleic acid encoding a IRE1 protein that
CC modulates unfolded protein response and has at least 60% similarity to a
CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
CC having unfolded protein response modulating activity and having greater
CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
CC heterologous nucleic acid encoding a protein having unfolded protein
CC response modulating activity and a heterologous nucleic acid encoding a
CC protein of interest to be secreted. The methods and compositions of
CC genetically manipulating cells to have an elevated unfolded protein
CC response (UPR) resulting in an increased capacity to produce secreted
CC proteins, are useful in e.g. production of therapeutic or industrial
CC enzymes. This is the amino acid sequence of Trichoderma reesei HAC1
CC chaperone and foldase DNA binding domain.

XX Sequence 451 AA;

Query Match 100.0%; Score 323; DB 8; Length 451;
Best Local Similarity 100.0%; Pred. No. 1.3e-28;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKKPVKKRSWGQVLPEPKTNLPKRKRAKTEDEKEQRRVERVLRNRAAQSRRKRLEV 60

DB 84 EKKPVKKRSWGQVLPEPKTNLPKRKRAKTEDEKEQRRVERVLRNRAAQSRRKRLEV 143

QY 61 EALE 64

DB 144 EALE 147

RESULT 6

ABO43144

ID ABO43144 standard; protein; 409 AA.

XX ABO43144;

DT 23-SEP-2003 (first entry)

DE A. thaliana disease tolerance transcription factor, G1034.

XX Plant; transcription factor; disease resistance; transgenic;
KW plant breeding; pathogens resistance; pests; resistance.

OS Arabidopsis thaliana.

PN US2003046723-A1.

XX 06-MAR-2003.

PF 22-MAR-2000; 2000US-00533029.

PR 22-MAR-2000; 2000US-00533029.

XX (HEAR/) HEARD J.

PA (BROU/) BROUN P.

PA (RIEC/) RIECHMANN J L.

PA (KEDD/) KEDDIE J.

PA (PINE/) PINEDA O.

PA (ADAM/) ADAM L.

PA (SAMA/) SAMAHA R.

PA (ZHAN/) ZHANG J.

PA (YUGG/) YU G.

PA (RATC/) RATCLIFFE O.

PA (PILG/) PILGRIM M.

PA (JIAN/) JIANG C.

PA (REUB/) REUBER L.

XX Heard J, Broun P, Riechmann JL, Keddie J, Pineda O, Adam L;

PI Samaha R, Zhang J, Yu G, Ratcliffe O, Pilgrim M, Jiang C, Reuber L;

XX WPI; 2003-521768/49.

DR N-PSDB; ACD98410.

XX

PT New transgenic plants comprising a recombinant gene that alters the
PT plant's disease tolerance or resistance, useful in plant breeding, e.g.
PT for generating plants with improved tolerance or resistance to diseases,
PT pests or pathogens.

XX Claim 1; Page 100-101; 124pp; English.

XX The invention relates to a transgenic plant, comprising a recombinant
CC polynucleotide that alters the plant's disease tolerance or resistance
CC when compared with the same trait of another plant lacking the
CC recombinant polynucleotide. The recombinant polynucleotide comprises a
CC nucleotide sequence, which encodes a polypeptide comprising at least 6
CC consecutive amino acids of any of 56 transcription factor proteins
CC appearing as ABO43093-ABO43148. Also included are altering the disease
CC tolerance or resistance of a plant (by: (a) transforming a plant with the
CC recombinant polynucleotide; (b) selecting the transformed plants; and (c)
CC identifying a transformed plant with an altered disease tolerance or
CC resistance), altering the expression levels of at least one gene in a
CC plant by transforming the plant with the recombinant polynucleotide,
CC altering a plant's trait (comprising: (a) providing a database sequence;
CC (b) comparing the database sequence with the polypeptide or
CC polynucleotide cited above; (c) selecting a database sequence that meets
CC the selected sequence criteria; and (d) transforming the selected
CC database sequence in the plant) and altering a plant's trait (comprising:
CC (a) providing a test polynucleotide; (b) hybridising the test
CC polynucleotide at low stringency with the recombinant polynucleotide
CC cited above; and (c) transforming the hybridising test polynucleotide
CC a plant to alter a trait of the plant. The transgenic plant is useful in
CC plant breeding, particularly for generating plants with improved
CC tolerance or resistance to diseases. The plants have commercial utility
CC for increasing tolerance or resistance to pathogens and pests. The
CC present sequence is an Arabidopsis thaliana transcription factor of the
CC invention

XX Sequence 409 AA;

Query Match 95.0%; Score 307; DB 7; Length 409;

Best Local Similarity 93.8%; Pred. No. 7.9e-27;

Matches 60; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKKPVKKRSWGQVLPEPKTNLPKRKRAKTEDEKEQRRVERVLRNRAAQSRRKRLEV 60

DB 66 EKKPVKKRSWGQVLPEPKTNLPKRKRAKTEDEKEQRRVERVLRNRAAQSRRKRQEV 125

QY 61 EALE 64

DB 126 EALE 129

RESULT 7

ADB31925

ID ADB31925 standard; protein; 409 AA.

XX ADB31925;

DT 04-DEC-2003 (first entry)

DE Plant (A. thaliana) transcription factor polypeptide #79.

XX Plant; transcription factor; transgenic plant; transgenic; plant trait;
KW modified trait.

OS Arabidopsis thaliana.

PN US2003101481-A1.

XX 29-MAY-2003.

PF 15-NOV-2002; 2002US-00295403.

XX 22-SEP-1998; 98US-0101349P.

PR 06-OCT-1998; 98US-0103312P.

PR 17-NOV-1998; 98US-0108734P.

PR 22-DEC-1998; 98US-0113409P.
 PR 13-SEP-1999; 99US-00394519.
 XX (ZHAN/) ZHANG J.
 PA (FROM/) FROMM M.
 PA (HEAR/) HEARD J.
 PA (RIEC/) RIECHMANN J L.
 PA (ADAM/) ADAM L.
 PA (BROU/) BROUN P.
 PA (PINE/) PINEDA O.
 PA (REUB/) REUBER L.
 PA (KEDD/) KEDDIE J.
 PA (YUGG/) YU G.
 PA (JIANG/) JIANG C.
 XX Zhang J, Fromm M, Heard J, Riechmann JL, Adam L, Broun P,
 PI Pineda O, Reuber L, Keddle J, Yu G, Jiang C;
 XX WPI; 2003-597572/56.
 DR N-PSDB; ADB31924.
 XX New isolated polypeptides and polynucleotide sequences, useful for
 PT screening a molecule to identify a molecule that modifies plant trait,
 PT and for producing plants with modified traits.
 PS Claim 9; SEQ ID NO 158; 17pp; English.
 XX The present invention relates to the isolation of plant (Arabidopsis
 CC thaliana) transcription factor polypeptide and polynucleotide sequences.
 CC Also disclosed are: an expression vector comprising the isolated
 CC polynucleotide, a host cell comprising the expression vector, a
 CC transgenic plant comprising the isolated polynucleotide, a transgenic
 CC plant ectopically expressing the isolated polynucleotide or polypeptide,
 CC a method for screening a molecule to identify a molecule that modifies a
 CC plant trait by placing the molecule in contact with the plant, and
 CC monitoring the effect of the molecule on the expressing or activity of
 CC the polypeptide or polynucleotide, and producing a transgenic plant
 CC having a modified trait by ectopically expressing the isolated
 CC polypeptide and selecting a plant with the modified trait. The
 CC polypeptides, polynucleotides and methods are useful for screening a
 CC molecule to identify a molecule that modifies plant trait, and for
 CC producing plants with modified traits. The present sequence represents a
 CC plant transcription factor polypeptide of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site at seqdata.uspto.gov.
 XX Sequence 409 AA;
 SQ
 Query Match 95.0%; Score 307; DB 7; Length 409;
 Best Local Similarity 93.8%; Pred. No. 7.9e-27;
 Matches 60; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 EKKPKKRSWGQVLPEPTNLPPRKRAKTEDEKQRRVERVLRNRAAQSRRKRLEV 60
 DB 66 EKKPKKRSWGQQLPEPTNLPPRKRAKTEDEKQRRVERVLRNRAAQSRRKRQEV 125
 QY 61 EALE 64
 DB 126 EALE 129
 RESULT 8
 ADO02271
 ID ADO02271 standard; protein; 409 AA.
 XX
 AC ADO02271;
 XX
 DT 01-JUL-2004 (first entry)
 XX
 DE Thalecress transcription factor protein #342.
 XX
 KW Thalecress; transcription factor; plant; transgenic; abiotic stress;

KW cold tolerance; heat tolerance; drought; osmotic stress;
 KW phosphate limitation; potassium limitation; nitrogen limitation;
 KW hormone sensitivity; disease resistance; sugar sensing; seed germination;
 KW flowering; inflorescence architectural change;
 KW meristem cell differentiation; phyllotaxy; apical dominance;
 KW trichome development; seed development; premature senescence;
 KW delayed senescence; lethality; necrosis; plant size; leaf morphology;
 KW seed morphology; secondary metabolism; light response; shade avoidance.
 XX Arabidopsis thaliana.
 OS
 XX US2004045049-A1.
 PN
 XX
 XX 04-MAR-2004.
 PD
 XX
 PF 10-APR-2003; 2003US-00412699.
 XX
 PR 13-SEP-1999; 99US-00394519.
 PR 21-JAN-2000; 2000US-00489376.
 PR 17-FEB-2000; 2000US-00506720.
 PR 22-MAR-2000; 2000US-00532591.
 PR 22-MAR-2000; 2000US-00533029.
 PR 22-MAR-2000; 2000US-00533030.
 PR 22-MAR-2000; 2000US-00533392.
 PR 22-MAR-2000; 2000US-00533648.
 PR 06-APR-2000; 2000WO-US009448.
 PR 16-NOV-2000; 2000US-00713994.
 PR 17-MAR-2001; 2001US-00819142.
 PR 17-APR-2001; 2001US-00837444.
 PR 30-JAN-2002; 2002US-00958131.
 PR 14-JUN-2002; 2002US-00171468.
 PR 09-AUG-2002; 2002US-00225066.
 PR 09-AUG-2002; 2002US-00225067.
 PR 09-AUG-2002; 2002US-00225068.
 PR 17-DEC-2002; 2002US-0434166P.
 PR 25-FEB-2003; 2003US-00374780.
 XX (ZHAN/) ZHANG J.
 PA (FROM/) FROMM M.
 PA (HEAR/) HEARD J.
 PA (RIEC/) RIECHMANN J L.
 PA (ADAM/) ADAM L J.
 PA (BROU/) BROUN P E.
 PA (PINE/) PINEDA O.
 PA (REUB/) REUBER T L.
 PA (KEDD/) KEDDIE J S.
 PA (YUGG/) YU G.
 PA (JIANG/) JIANG C.
 PA (SAMA/) SAMAHA R S.
 PA (PILG/) PILGRIM M L.
 PA (CREE/) CREELMAN R A.
 PA (DUBE/) DUBELL A N.
 PA (RATC/) RATCLIFFE O.
 PA (KUMI/) KUMIMOTO R.
 PA (SHER/) SHERMAN B K.
 XX Zhang J, Fromm ME, Heard JE, Riechmann JL, Adam LJ, Broun PE;
 PI Pineda O, Reuber TL, Keddle JS, Yu G, Jiang C, Samaha RS;
 PI Pilgrim ML, Creelman RA, Dubell AN, Ratcliffe O, Kumimoto R;
 PI Sherman BK;
 XX WPI; 2004-225755/21.
 DR N-PSDB; ADO02270.
 DR
 XX New transgenic plant, useful in developing phenotypes with altered or
 PT improved characteristics or traits.
 PT
 XX Claim 1; SEQ ID NO 684; 213pp; English.
 PS
 XX The invention relates to a transgenic plant comprises a recombinant
 CC polynucleotide having a polynucleotide sequence or its complementary
 CC sequence comprising a sequence encoding a polypeptide that initiates
 CC transcription (i.e. a transcription factor) from Arabidopsis, Soybean,

KW cellulase; carbohydrase; industry; purification; haca protein.
 XX Aspergillus niger.
 OS
 XX
 PN US2001034045-A1.
 XX
 PD 25-OCT-2001.
 XX
 XX
 PF 23-MAR-2001; 2001US-00816277.
 XX
 PR 24-MAR-2000; 2000US-00534692.
 XX
 XX (GEMV) GENENCOR INT INC.
 PA
 XX Penttala ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
 PI WPI; 2002-033728/04.
 DR
 XX
 XX Increasing secretion of heterologous proteins e.g. lipase and cellulase
 PT in eukaryotic cells useful in industry to increase production and
 PT facilitate purification, by inducing an elevated unfolded protein
 PT response.
 XX
 XX Claim 39; Fig 28; 56pp; English.
 PS
 XX The present invention relates to methods for increasing the secretion of
 CC heterologous protein in eukaryotic cells by inducing an elevated unfolded
 CC protein response (UPR). The method involves inducing the elevated UPR by
 CC increasing the presence of proteins such as HAC1, HACA, PTC2 or IRE1 in
 CC cells. The method and sequences are useful for increasing the secretion
 CC of heterologous proteins (e.g. lipase, cellulase, carbohydrase) in
 CC eukaryotic cells useful in industry to increase protein yields and to
 CC facilitate purification. The present sequence is Aspergillus niger var.
 CC awamori haca protein
 XX
 XX Sequence 342 AA;
 SQ
 Query Match 90.4%; Score 292; DB 5; Length 342;
 Best Local Similarity 89.1%; Pred. No. 3.5e-25;
 Matches 57; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 1 EKKPVKKKSWGQVLPEPTNLPPRKRAKTEDEKEQRRVERVLRNRAAQSRRKRLEV 60
 DB 46 EKKPVKKKSWGQELPVPKTNLPFRKRAKTEDEKEQRRVERVLRNRAAQTSRKRLEM 105
 QY 61 EALE 64
 DB 106 EKLE 109
 RESULT 11
 ADS12818
 ID ADS12818 standard; protein; 342 AA.
 XX
 AC ADS12818;
 XX
 DT 16-DEC-2004 (first entry)
 XX
 XX Aspergillus nidulans haca polypeptide seqid 19.
 DE
 DE unfolded protein response; UPR; HAC1; PTC2;
 KW unfolded protein response modulator; enzyme production; haca; chaperone;
 KW foldase.
 XX
 OS Emericella nidulans.
 XX
 PN US2004186070-A1.
 XX
 PD 23-SEP-2004.
 XX
 PF 15-SEP-2003; 2003US-00663450.
 XX
 XX 24-MAR-2000; 2000US-00534692.
 PR

PR 23-MAR-2001; 2001US-00816277.
 XX
 PA (GEMV) GENENCOR INT INC.
 XX
 PI Penttala ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
 XX
 DR WPI; 2004-707924/69.
 XX
 XX Increasing the secretion of a heterologous protein, such as a therapeutic
 PT or an industrial enzyme, in genetically modified eukaryotic cells by
 PT inducing an elevated unfolded protein response (UPR).
 XX
 PS Example 12; SEQ ID NO 19; 83pp; English.
 XX
 XX The invention describes a method of increasing the secretion of a
 CC heterologous protein in a eukaryotic cell, comprising inducing an
 CC elevated unfolded protein response (UPR). Also described are: an isolated
 CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
 CC unfolded protein response and has less than 50% similarity to yeast HAC1
 CC protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
 CC protein induces unfolded protein response and where the HAC1 protein
 CC comprises a DNA binding region that has greater than 70% similarity to
 CC the DNA binding region of filamentous fungi HAC1 protein; a protein
 CC having unfolded protein response inducing activity and having greater
 CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
 CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
 CC respectively) as given in the specification; a protein having an amino
 CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
 CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
 CC response and has at least 70% similarity to a fully defined amino acid
 CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
 CC specification; an isolated nucleic acid encoding a IRE1 protein that
 CC modulates unfolded protein response and has at least 60% similarity to a
 CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
 CC having unfolded protein response modulating activity and having greater
 CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
 CC heterologous nucleic acid encoding a protein having unfolded protein
 CC response modulating activity and a heterologous nucleic acid encoding a
 CC protein of interest to be secreted. The methods and compositions of
 CC genetically manipulating cells to have an elevated unfolded protein
 CC response (UPR) resulting in an increased capacity to produce secreted
 CC proteins, are useful in e.g. production of therapeutic or industrial
 CC enzymes. This is the amino acid sequence of an Aspergillus nidulans haca
 CC chaperone and foldase polypeptide.
 XX
 XX Sequence 342 AA;
 SQ
 Query Match 90.4%; Score 292; DB 8; Length 342;
 Best Local Similarity 89.1%; Pred. No. 3.5e-25;
 Matches 57; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 1 EKKPVKKKSWGQVLPEPTNLPPRKRAKTEDEKEQRRVERVLRNRAAQSRRKRLEV 60
 DB 46 EKKPVKKKSWGQELPVPKTNLPFRKRAKTEDEKEQRRVERVLRNRAAQTSRKRLEM 105
 QY 61 EALE 64
 DB 106 EKLE 109
 RESULT 12
 AAE15379
 ID AAE15379 standard; protein; 386 AA.
 XX
 AC AAE15379;
 XX
 DT 07-MAR-2002 (first entry)
 XX
 XX Aspergillus niger var. awamori haca protein #1.
 DE
 DE Heterologous protein secretion; unfolded protein response; UPR; lipase;
 KW cellulase; carbohydrase; industry; purification; haca protein.
 XX

```
OS Aspergillus niger.
PN US2001034045-A1.
XX
PD 25-OCT-2001.
XX
PF 23-MAR-2001; 2001US-00816277.
XX
PR 24-MAR-2000; 2000US-00534692.
XX
PA (GEMV ) GENENCOR INT INC.
XX
PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX
DR WPI; 2002-033728/04.
DR N-PSDB; AAD24601.
XX
PT Increasing secretion of heterologous proteins e.g. lipase and cellulase
PT in eukaryotic cells useful in industry to increase production and
PT facilitate purification, by inducing an elevated unfolded protein
PT response.
XX
PS Claim 39; Fig 28; 56pp; English.
XX
CC The present invention relates to methods for increasing the secretion of
CC heterologous protein in eukaryotic cells by inducing an elevated unfolded
CC protein response (UPR). The method involves inducing the elevated UPR by
CC increasing the presence of proteins such as HAC1, HAC2, PTC2 or IRE1 in
CC cells. The method and sequences are useful for increasing the secretion
CC of heterologous proteins (e.g. lipase, cellulase, chaperone) in
CC eukaryotic cells useful in industry to increase protein yields and to
CC facilitate purification. The present sequence is Aspergillus niger var.
CC awamori hacA protein
XX
SQ Sequence 386 AA;
Query Match 90.4%; Score 292; DB 5; Length 386;
Best Local Similarity 89.1%; Pred. No. 4e-25;
Matches 57; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 1 EKKPVKGRKSWGQVLPEPKTNLPKRKRAKTEDEKQRRVERVLNRRAAQSRRKRLEV 60
DB 90 EKKPVKGRKSWGQELVPKTNLPKRKRAKTEDEKQRRVERVLNRRAAQSRRKRLEM 149
QY 61 EALE 64
DB 150 EKLE 153
RESULT 13
ADSI2815
ID ADSI2815 standard; protein; 386 AA.
XX
AC ADSI2815;
XX
DT 16-DEC-2004 (first entry)
XX
DE Aspergillus nidulans hacA chaperone and foldase #3.
XX
KW unfolded protein response; UPR; HAC1; PTC2;
KW unfolded protein response modulator; enzyme production; hacA; chaperone;
KW foldase; enzyme.
XX
OS Emericella nidulans.
XX
PN US2004186070-A1.
XX
PD 23-SEP-2004.
XX
PF 15-SEP-2003; 2003US-00663450.
XX
PR 24-MAR-2000; 2000US-00534692.
PR 23-MAR-2001; 2001US-00816277.
```

```
XX (GEMV ) GENENCOR INT INC.
XX
XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX
XX WPI; 2004-707924/69.
XX N-PSDB; ADSI2814.
XX
XX Increasing the secretion of a heterologous protein, such as a therapeutic
XX or an industrial enzyme, in genetically modified eukaryotic cells by
XX inducing an elevated unfolded protein response (UPR).
XX
XX Example 12; SEQ ID NO 16; 83pp; English.
XX
XX The invention describes a method of increasing the secretion of a
XX heterologous protein in a eukaryotic cell, comprising inducing an
XX elevated unfolded protein response (UPR). Also described are: an isolated
XX nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
XX unfolded protein response and has less than 50% similarity to yeast HAC1
XX protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
XX protein induces unfolded protein response and where the HAC1 protein
XX comprises a DNA binding region that has greater than 70% similarity to
XX the DNA binding region of filamentous fungi HAC1 protein; a protein
XX having unfolded protein response inducing activity and having greater
XX than 70% similarity to a fully defined amino acid sequence of 451, 349 or
XX 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
XX respectively) as given in the specification; a protein having an amino
XX acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
XX nucleic acid encoding a PTC2 protein that modulates unfolded protein
XX response and has at least 70% similarity to a fully defined amino acid
XX sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
XX specification; an isolated nucleic acid encoding a IRE1 protein that
XX modulates unfolded protein response and has at least 60% similarity to a
XX fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
XX having unfolded protein response modulating activity and having greater
XX than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
XX heterologous nucleic acid encoding a protein having unfolded protein
XX response modulating activity and a heterologous nucleic acid encoding a
XX protein of interest to be secreted. The methods and compositions of
XX genetically manipulating cells to have an elevated unfolded protein
XX response (UPR) resulting in an increased capacity to produce secreted
XX proteins, are useful in e.g. production of therapeutic or industrial
XX enzymes. This is the amino acid sequence of Aspergillus nidulans HAC1
XX chaperone and foldase.
XX
XX Sequence 386 AA;
Query Match 90.4%; Score 292; DB 8; Length 386;
Best Local Similarity 89.1%; Pred. No. 4e-25;
Matches 57; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 1 EKKPVKGRKSWGQVLPEPKTNLPKRKRAKTEDEKQRRVERVLNRRAAQSRRKRLEV 60
DB 90 EKKPVKGRKSWGQELVPKTNLPKRKRAKTEDEKQRRVERVLNRRAAQSRRKRLEM 149
QY 61 EALE 64
DB 150 EKLE 153
RESULT 14
AAEI5374
ID AAEI5374 standard; protein; 64 AA.
XX
XX AAEI5374;
XX
XX 29-AUG-2003 (revised)
XX 07-MAR-2002 (first entry)
XX
XX Aspergillus nidulans hacA protein DNA binding domain.
XX
XX Heterologous protein secretion; unfolded protein response; UPR; lipase;
XX cellulase; carbohydrase; industry; purification; DNA binding domain;
KW
```

```

KW haca protein.
XX
OS Emericella nidulans.
XX
PN US2001034045-A1.
XX
PD 25-OCT-2001.
XX
PP 23-MAR-2001; 2001US-00816277.
XX
PR 24-MAR-2000; 2000US-00534692.
XX
PA (GEMV ) GENENCOR INT INC.
XX
PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX WPI; 2002-033728/04.
XX
XX Increasing secretion of heterologous proteins e.g. lipase and cellulase
XX in eukaryotic cells useful in industry to increase production and
XX facilitate purification, by inducing an elevated unfolded protein
XX response.
XX
XX Example 3; Fig 10; 56pp; English.
XX
XX The present invention relates to methods for increasing the secretion of
XX heterologous protein in eukaryotic cells by inducing an elevated unfolded
XX protein response (UPR). The method involves inducing the elevated UPR by
XX increasing the presence of proteins such as HAC1, HACA, PTC2 or IRE1 in
XX cells. The method and sequences are useful for increasing the secretion
XX of heterologous proteins (e.g. lipase, cellulase, carboxydase) in
XX eukaryotic cells useful in industry to increase protein yields and to
XX facilitate purification. The present sequence is Aspergillus nidulans
XX haca protein DNA binding domain. (Updated on 29-AUG-2003 to standardise
XX OS field)
XX
XX Sequence 64 AA;
XX
Query Match 89.2%; Score 288; DB 5; Length 64;
Best Local Similarity 87.5%; Pred. No. 1.6e-25;
Matches 56; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 EKKPVKKRKSQGVLPKPTNLPPKRAKTEDEKEQRRVRLNRRRAAQSRRKRLEV 60
DB 1 EKKPAKRRKSQGVLPKPTNLPPKRAKTEDEKEQRRVRLNRRRAAQSRRKRLEV 60
OY 61 EALE 64
DB 61 EKLE 64

RESULT 15
ADSI2805
ID ADSI2805 standard; protein; 349 AA.
XX
XX ADSI2805;
AC
XX
XX 16-DEC-2004 (first entry)
DE
DE Aspergillus nidulans haca DNA binding domain.
XX
XX unfolded protein response; UPR; HAC1; PTC2;
KW unfolded protein response modulator; enzyme production; haca; chaperone;
KW foldase; DNA binding domain.
XX
XX Emericella nidulans.
OS
XX
XX US2004186070-A1.
PN
XX
XX 23-SEP-2004.
PD
XX
XX 15-SEP-2003; 2003US-00663450.
PF
XX

```

```

PR 24-MAR-2000; 2000US-00534692.
XX
XX 23-MAR-2001; 2001US-00816277.
XX
XX (GEMV ) GENENCOR INT INC.
XX
XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX WPI; 2004-707924/69.
XX
XX Increasing the secretion of a heterologous protein, such as a therapeutic
XX or an industrial enzyme, in genetically modified eukaryotic cells by
XX inducing an elevated unfolded protein response (UPR).
XX
XX Example 3; SEQ ID NO 6; 83pp; English.
XX
XX The invention describes a method of increasing the secretion of a
XX heterologous protein in a eukaryotic cell, comprising inducing an
XX elevated unfolded protein response (UPR). Also described are: an isolated
XX nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
XX unfolded protein response and has less than 50% similarity to yeast HAC1
XX protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
XX protein induces unfolded protein response and where the HAC1 protein
XX comprises a DNA binding region that has greater than 70% similarity to
XX the DNA binding region of filamentous fungi HAC1 protein; a protein
XX having unfolded protein response inducing activity and having greater
XX than 70% similarity to a fully defined amino acid sequence of 451, 349 or
XX 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
XX respectively) as given in the specification; a protein having an amino
XX acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
XX nucleic acid encoding a PTC2 protein that modulates unfolded protein
XX response and has at least 70% similarity to a fully defined amino acid
XX sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
XX specification; an isolated nucleic acid encoding a IRE1 protein that
XX modulates unfolded protein response and has at least 60% similarity to a
XX fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
XX having unfolded protein response modulating activity and having greater
XX than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
XX heterologous nucleic acid encoding a protein having unfolded protein
XX response modulating activity and a heterologous nucleic acid encoding a
XX protein of interest to be secreted. The methods and compositions of
XX genetically manipulating cells to have an elevated unfolded protein
XX response (UPR) resulting in an increased capacity to produce secreted
XX proteins, are useful in e.g. production of therapeutic or industrial
XX enzymes. This is the amino acid sequence of Aspergillus nidulans HAC1
XX chaperone and foldase DNA binding domain.
XX
XX Sequence 349 AA;
XX
Query Match 89.2%; Score 288; DB 8; Length 349;
Best Local Similarity 87.5%; Pred. No. 1e-24;
Matches 56; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 EKKPVKKRKSQGVLPKPTNLPPKRAKTEDEKEQRRVRLNRRRAAQSRRKRLEV 60
DB 52 EKKPAKRRKSQGVLPKPTNLPPKRAKTEDEKEQRRVRLNRRRAAQSRRKRLEV 111
OY 61 EALE 64
DB 112 EKLE 115

Search completed: November 22, 2005, 15:43:30
Job time : 161.secs

```

THIS PAGE BLANK (USPTO)